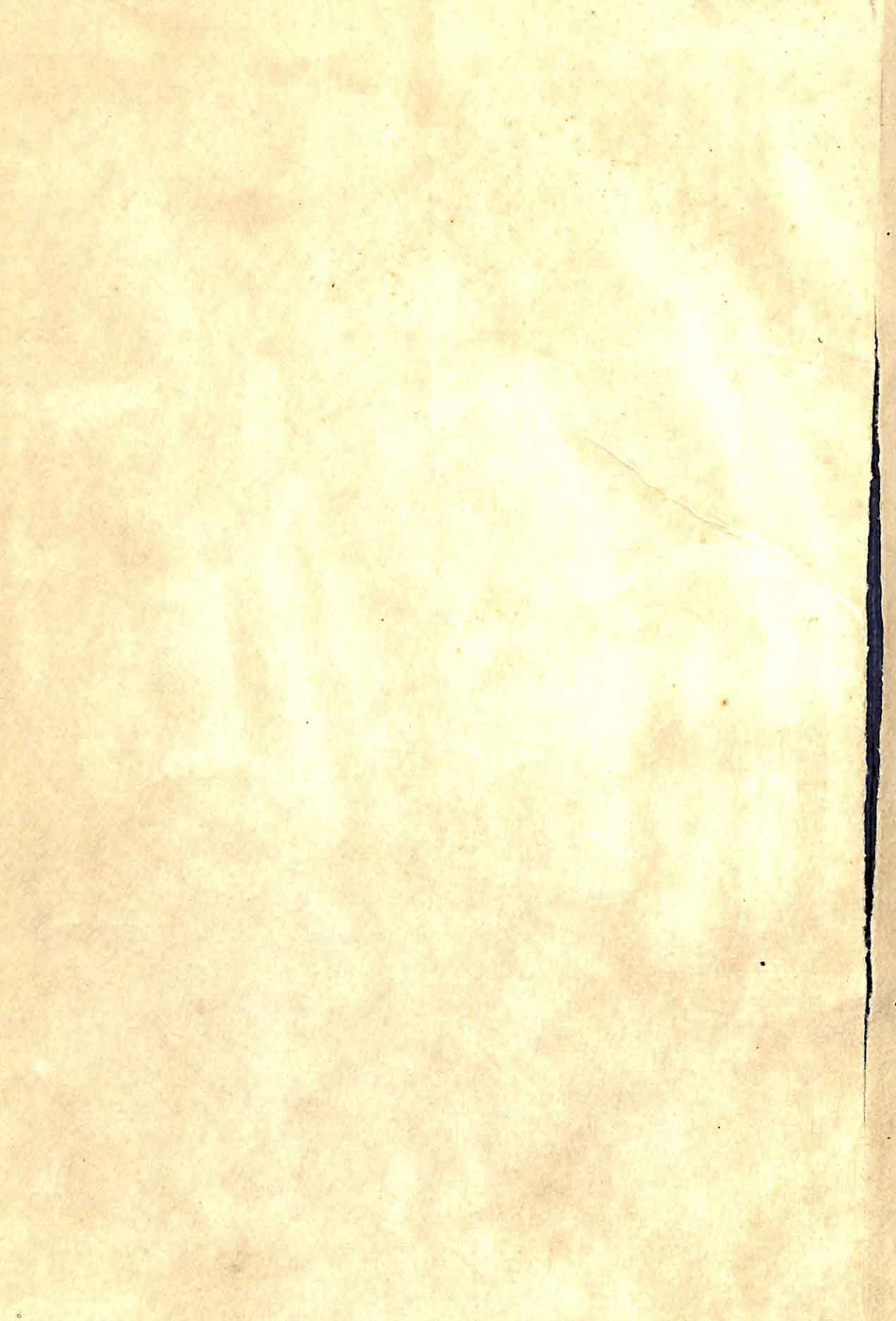




240  
27.8.20











# THE BRITISH JOURNAL OF SOCIAL AND CLINICAL PSYCHOLOGY

EDITED BY  
MICHAEL ARGYLE & JACK TIZARD

WITH THE ASSISTANCE OF

D. RUSSELL DAVIS  
CECILY de MONCHAUX  
H. J. EYSENCK  
G. A. FOULDS  
MAX HAMILTON

A. G. HAMMER  
D. W. HARDING  
ALASTAIR HERON  
HILDE T. HIMMELWEIT  
MARIE JAHODA

A. R. LURIA  
A. E. MAXWELL  
T. M. NEWCOMB  
R. R. SEARS  
E. J. SHOBN

**Volume 3, 1964**

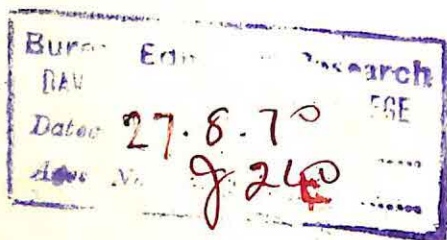


**CAMBRIDGE**

Published for the British Psychological Society  
AT THE UNIVERSITY PRESS

1964

PUBLISHED BY  
THE SYNDICS OF THE CAMBRIDGE UNIVERSITY PRESS  
Bentley House, 200 Euston Road, London N.W.1.  
American Branch: 32 East 57th Street, New York 22, N.Y.  
West African Office: P.O. Box 33, Ibadan, Nigeria



*Printed in Great Britain by Spottiswoode, Ballantyne & Co. Ltd.  
London and Colchester*



# CONTENTS OF VOLUME 3

## PART I. FEBRUARY 1964

COSTELLO, C. G. The Effects of Depressant and Stimulant Drugs on the Relationship between Reaction Time and Stimulus Light Intensity	page 1
TIZARD, BARBARA and MARGERISON, J. H. Psychological Functions during Wave-Spike Discharge	6
INGHAM, J. G. Cross-masking in Neurotic Patients	16
BROWN, ROY I. The Effect of Visual Distraction on Perception in Subjects of Subnormal Intelligence	20
O'CONNOR, N. and HERMELIN, BEATE. Measures of Distance and Motility in Psychotic Children and Severely Subnormal Controls	29
FOULDES, G. A. and OWEN, ANNA. Speed and Accuracy on Mazes in Relation to Diagnosis and Personality	34
POPPLETON, PAMELA K. and PILKINGTON, G. W. A Comparison of Four Methods of Scoring an Attitude Scale in Relation to its Reliability and Validity	36
TAJFEL, HENRI and WILKES, A. L. Salience of Attributes and Commitment to Extreme Judgments in the Perception of People	40
GIBSON, H. B. A Slang Vocabulary Test as an Indication of Delinquent Association	50
LAMBERT, WALLACE E. Cultural Comparison of Boy's Occupational Aspirations	56
VAUGHAN, GRAHAM M. The Effect of the Ethnic Grouping of the Experimenter upon Children's Responses to Tests of an Ethnic Nature	66
JONCKHEERE, A. R. Critical Notice	71
Book Reviews	74

## PART 2. JUNE 1964

DENTLER, ROBERT A. and MACKLER, BERNARD. Effects on Sociometric Status of Institutional Pressure to Adjust Among Retarded Children	page 81
PRITCHATT, DERRICK. A Note on the Interpretation of the Guess Who Test in the Study of Sociometric Choice Behaviour	90
KRIEGER, MARGERY H. A Control for Social Desirability in a Semantic Differential	94
SMITH, PETER B. Attitude Changes Associated with Training in Human Relations	104
AL-ISSA, IHSAN. The Effect of Attitudinal Factors on the Relationship Between Conditioning and Personality	113
RUTTER, MICHAEL. Intelligence and Childhood Psychiatric Disorder	120

MERSKEY, H. and SPEAR, F. G. The Reliability of the Pressure Algometer	130
CROOKES, T. G. and HUTT, S. J. A note on the MPI: A Response Tendency in Neurotics Related to Intelligence	137
MAY, A. E., LYND, B. R. and LEY, P. The Effects of Distributed Practice on the Learning Performance of Brain Damaged Patients	139
EYSENCK, SYBIL B. G. and EYSENCK, HANS J. The Personality of Judges as a Factor in the Validity of Their Judgments of Extraversion-Introversion	141
GOLDBERG, LEWIS R. Intra-Individual Variability in the MMPI-CPI Common Item Pool	145
BOLARDOS, A. C. Validation of the Maudsley Personality Inventory in Chile	148
JALOTA, S. Some Data on the Maudsley Personality Inventory in Panjabi	148
BARBER, THEODORE XENOPHON and CALVERLEY, DAVID SMITH. Hypnotizability, Suggestibility, and Personality: IV. A Study with the Leary Interpersonal Check List	149
GIBSON, H. B. Neuroticism, Extraversion and Hypnotizability: A Reply	151
Book Reviews	153

### PART 3. OCTOBER 1964

ORME, J. E., LEE, D. and SMITH M. R. Psychological Assessments of Brain Damage and Intellectual Impairment in Psychiatric Patients	page 161
CORAH, NORMAN L. Neuroticism and Extraversion in the MMPI: Empirical Validation and Exploration	168
DAVIDSON, M. A. and HUTT, CORINNE. A Study of 500 Oxford Student Psychiatric Patients	175
HAWKS, D. V. The Clinical Usefulness of some tests of Over-inclusive Thinking in Psychiatric Patients	186
KELLMER PRINGLE, M. L. and EDWARDS, J. B. Some Moral Concepts and Judgments of Junior School Children	196
MARTIN, JOHN. Acquiescence—Measurement and Theory	216
COLQUHOUN, W. P. and CORCORAN, D. W. J. The Effects of Time of Day and Social Isolation on the Relationship Between Temperament and Performance	226
Book Reviews	232



## INDEX TO AUTHORS

AL-ISSA, IHSAN . . . . .	113	KELLMER PRINGLE, M. L. . . . .	196
BARBER, THEODORE XENOPHON . . . . .	149	KRIEGER, MARGERY H. . . . .	94
BOLARDOS, A. C. . . . .	148	LAMBERT, WALLACE E. . . . .	56
BROWN, ROY, I . . . . .	20	LEE, D. . . . .	161
CALVERLEY, DAVID SMITH . . . . .	149	LEY, P. . . . .	139
COLQUHOUN, W. P. . . . .	226	LYND, B. R. . . . .	139
CORAH, NORMAN L. . . . .	168	MACKLER, BERNARD. . . . .	81
CORCORAN, D. W. J. . . . .	226	MARGERISON, J. A. . . . .	6
COSTELLO, C. G. . . . .	1	MARTIN, JOHN . . . . .	216
CROOKES, T. G. . . . .	137	MAY, A. E. . . . .	139
DAVIDSON, M. A. . . . .	175	MERSKEY, H. . . . .	130
DENTLAR, ROBERT A. . . . .	81	O'CONNOR, N. . . . .	29
EDWARDS, J. B. . . . .	196	ORME, J. E. . . . .	161
EYSENCK, HANS J. . . . .	141	OWEN, ANNA . . . . .	34
EYSENCK, SYBIL B. G. . . . .	141	PILKINGTON, G. W. . . . .	3.
FOULDES, G. A. . . . .	34	POPPLTON, PAMELA K. . . . .	36
GIBSON, . H. . . . .	50, 151	PRITCHATT, DERRICK . . . . .	90
GOLDBERG, LEWIS R. . . . .	145	RUTTER, MICHAEL . . . . .	120
HAWKS, D. V. . . . .	186	SMITH, M. R. . . . .	161
HERMELIN, BEATE . . . . .	29	SMITH, PETER, B. . . . .	104
HUTT, CORINNE . . . . .	175	SPEAR, F. G. . . . .	130
HUTT, S. J. . . . .	137	TAJFEL, HENRI . . . . .	40
INGHAM, J. G. . . . .	16	TIZARD, BARBARA . . . . .	6
JALOTA, S. . . . .	148	VAUGHAN, GRAHAM M. . . . .	66
JONCKHEERE, A. R. . . . .	71	WILKES, A. L. . . . .	40

## INDEX OF PUBLICATIONS REVIEWED

ALCOCK, THEODORA, <i>The Rorschach in Practice</i> 240	GRÜNHUT, M. <i>Probation and Mental Treatment</i> 238
ANSTAY, E. <i>Committees—How They Work and How to Work Them</i> 160	HAGEN, EVERETTE. <i>On the Theory of Social Change</i> 236
BANDURA, ALBERT and WALTERS, RICHARD H. <i>Social Learning and Personality Development</i> 234	HALMES, P. (ed.). <i>The Sociological Review Monograph No. 7: Sociological Studies in British University Education</i> 159
CAMERON, NORMAN. <i>Personality Development Psychopathology</i> 232	HALMOS, P. (ed.). <i>The Canford Families. A Study of Social Casework and Group Work</i> 77
COFER, CHARLES N. (ed.). <i>Verbal Learning and Verbal Behaviour</i> 77	HOLLANDER, E. P. & HUNT, R. G. (eds.). <i>Current Perspectives in Social Psychology: Readings with Commentary</i> 157
COOPER, J. B. & MCGAUGH, J. L. <i>Integrating Principles of Social Psychology</i> 236	HUMPHREY, G. (ed.). <i>Psychology Through Experiment</i> 76
DONALDSON, MARGARET, WITHRINGTON, DONALD & DUTHIE, JOHN. <i>A Study of Children's Thinking</i> 160	HUNTER, RICHARD & MACALPINE, IDA. <i>Three Hundred Years of Psychiatry 1535-1860</i> 78
DULANY, D. E., DEVALOIS, R. L., BEARDSLEE, D. C. & WINTERBOTTOM, M. R. (eds.). <i>Contribution to Modern Psychology: Selected Readings in General Psychology</i> 240	JONES, HOWARD, <i>Alcoholic Addiction. A Psycho-social Approach to Abnormal Drinking</i> 239
FESTINGER, L., SCHACHTER, S. & BACK, K. <i>Social Pressures in Informal Groups: a Study of Human Factors in Housing</i> 238	KRAEPELIN, EMIL. <i>One hundred Years of Psychiatry</i> 78
FISCHER, H. <i>Gruppenstruktur und Gruppenleistung</i> 80	KRECH, D., CRUTCHFIELD, R. S. & BALLACHEY, E. L. <i>Individual in Society</i> 153
FREEMAN, HUGH & FARNDAL, JAMES (eds.). <i>Friends in the Mental Health Services</i> 237	LUCE, R. D., BUSH, R. R. & GALANTER, E. (eds.). <i>Readings in Mathematical Psychology</i> (vol. 1) 155

- LURIA, A. R. (ed.). *The Mentally Retarded Child* 154
- MALAN, D. H. *A Study of Brief Psychotherapy* 239
- MASON, STELLA E. (ed.) *Signs, Signals and Symbols* 80
- MEDNICK, MARTA T. & MEDNICK, SARNOFF A. (eds.). *Research in Personality* 75
- MOWBRAY, R. M. & RODGER, T. FERGUSON. *Psychology in Relation to Medicine* 156
- PALERMO, DAVID S. & LIPSITT, LEWIS, P. (eds.). *Research Readings in Child Psychology* 158
- PICKFORD, R. W. (et al.). *Pickford Projective Pictures* 233
- ROUCEK, J. (ed.). *The Unusual Child* 79
- RUBENOWITZ, SIGVARD. *Emotional Flexibility—Rigidity as a Comprehensive Dimension of Mind* 238
- SPENCER, JOHN, TUXFORD, JOY & DENNIS, NORMAN. *Stress and Release in an Urban Estate—A Study in Action Research* 235
- SPINKS, G. STEPHENS. *Psychology and Religion* 156
- SUNDBERG, N. D. & TYLER, L. E. *Clinical Psychology: An Introduction to Research and Practice* 232
- TRIST, E. L., HIGGIN, G. W., MURRAY, H. & POLLOCK, A. B. *Organizational Choice* 158
- WHITING, BEATRICE B. (ed.). *Six Cultures: Studies in Child Rearing* 153



## The Effects of Depressant and Stimulant Drugs on the Relationship between Reaction Time and Stimulus Light Intensity

By C. G. COSTELLO

*Regina Psychological Laboratory, Munroe Wing, Regina, Saskatchewan, Canada*

A paradoxical increase in RT with increase in visual stimulus intensity has been found by previous workers using schizophrenic Ss. Their results had been interpreted in terms of Pavlovian theory. It was predicted on the basis of Pavlovian theory and previous results with schizophrenic Ss that both 10 mg D-amphetamine sulphate by increasing the normal Ss level of excitation and 194 mg sodium seconal by decreasing his level of excitation would result in the paradoxical increase in RT with an increase in stimulus intensity.

Five groups of eight Ss each were tested after administration of one of the following treatments: 10 mg D-amphetamine sulphate; 194 mg sodium seconal; 800 mg meprobamate; placebo tablets; no drug. Their RTs to lights of eight intensities increasing from 16 ft candles to 1500 ft candles were determined. The predictions were not confirmed.

The sodium seconal, meprobamate and placebo groups produced RTs significantly different from those of the D-amphetamine and no drug groups which did not differ in their mean RTs.

Implications of the results in terms of the difference between cortical and subcortical excitation are discussed.

### INTRODUCTION

Venables & Tizard (1956*b*) reported that in 22 out of 24 chronic, non-paranoid schizophrenics, RTs to the brighter light stimuli were longer than those to weaker stimuli. These results were in keeping with those found in a pilot study (Venables & Tizard, 1956*a*). This increase in the RTs of schizophrenics when reacting to lights of higher intensity is in contrast to the established finding in normal Ss of a decrease in RT with an increase in stimulus intensity (Berger, 1886; Cattell, 1886; Huil, 1949).

This paradoxical effect of an increase in RT with an increase in stimulus intensity beyond an optimum point only occurred on initial testing in the study by Venables and Tizard. It did not occur on a second occasion of testing. The authors write, 'A possible explanation is that on the first occasion, in an unfamiliar situation, Ss were apprehensive and in a state of considerable emotional excitation. The excitation resulting from the stimulus lights, superimposed on this general excitation became ultramarginal in strength and resulted in the paradoxical effect observed. On the second occasion of testing, Ss were more familiar with the situation, and the total excitation did not reach ultramarginal strength. This dependence of the paradoxical effect on the Ss level of excitation is in line with Pavlov's views on the phenomenon . . . and it is predictable from his theory that after extensive testing when the general level of excitation of the S is very low because of boredom or fatigue, the paradoxical effect reappears. If subsequent experimentation should



prove this to be the case, the second occasion of testing could probably be regarded as an intermediate state between two phasic states.'

In general it is suggested by Venables and Tizard that the paradoxical effect will occur in states of very strong excitation or very weak excitation.

The purpose of the present study was to investigate the effects of altering the *Ss* level of excitation, by the administration of drugs, on his RT to lights of increasing intensity.

Specifically it was predicted: (1) That D-amphetamine sulphate, by increasing the level of excitation, would result in the paradoxical increase in RT with an increase in stimulus intensity. (2) That sodium seconal, by lowering the level of excitation, would result in the paradoxical increase in RT with an increase in stimulus intensity.

The effect of meprobamate, which acts as a mild depressant (Costello, 1963), was also investigated in order that a comparison could be made between two levels of lowered excitation—mild (meprobamate) and strong (sodium seconal).

### *Subjects*

Forty *Ss* were tested, 20 male and 20 female. Their ages ranged from 17 years to 23 years with a mean of 19.31 years. All the *Ss* were College students and were paid volunteers.

### *Apparatus and procedure*

The apparatus used was in all essentials the same as that used by Venables & Tizard (1956b).

The stimulus was an aperture 0.75 in. in diameter illuminated from the rear by a 150 watt bulb fed from a stabilized voltage source. The light source was set in a 2 ft. × 2 ft. matt black panel placed at an angle to the vertical so as to be at right angles to the *Ss* line of sight when seated at a distance of 2 ft from the panel. Variations in stimulus intensity were made by means of neutral density filters. The intensities of the stimuli were measured by means of a photometer. The values obtained are given in Table 1.

Table 1. *Visual stimulus intensities in foot candles*

Intensity	A	B	C	D	E	F	G	H
Foot Candles	16	60	135	275	450	600	800	1500

The group of 40 *Ss* was split randomly into five groups of eight *Ss* each. Each group received one of the following treatments: 10 mg D-amphetamine sulphate; 194 mg sodium seconal; 800 mg meprobamate; placebo tablets; no drug. All treatments were given 1½ h before testing.

Order of presentation of the eight intensities of stimulus light was varied in a balanced Latin-Square design for each of the subgroups of eight *Ss*. Twenty RTs were measured for each stimulus intensity though only responses 6–15 were used in the analysis to avoid warm-up and end-spurt effects.

A forewarning buzzer signal preceded by 3.0 sec the onset of the stimulus light. The *S* reacted to the stimulus light by lifting his finger. The light then went out and was prevented by a relay mechanism from immediately reappearing when *Ss* finger was replaced on the key to await the next presentation of the stimulus light. If *S* failed to respond, the light remained on for 3.0 sec. The cycle of buzzer and light was repeated once every 8 sec. Between each group of 20 responses, a 3 min rest period was allowed.

Measurement of each response was made by means of a tape recorder linked to a Lafayette stop clock No. 20225A accurate to the nearest hundredth of a second. This method of recording RTs has been described in detail elsewhere (Costello & Stephen, 1963).

## RESULTS

Table 2 shows the mean RTs for each of the five treatment groups. It can be seen from these data that, in general, the pattern in all cases is similar showing a slight fall in RT as stimulus intensity increases. Venables and Tizard noted that in their study there was only a slight fall in the RTs of their normal groups since at the intensities employed the slope of the curve of performance was approaching the asymptote.

Table 2. Mean RTs of five treatment groups ( $N = 8$  in each group), in 1/100 sec (decimal points removed)

Subjects	Stimulus intensity							
	A	B	C	D	E	F	G	H
No drug	30	29	29	29	27	28	28	28
Placebo	32	31	31	31	33	31	32	31
10 mg D-Amphetamine sulphate	30	29	31	29	27	28	28	28
800 mg Meprobamate	35	34	34	34	34	34	34	34
194 mg Sodium seconal	36	35	35	36	36	36	35	35

Though we were not primarily concerned in the present study with the overall effects of the drugs, it is noteworthy that mean response times over all intensities for the sodium seconal, meprobamate and placebo subgroups are significantly different from the mean response time of the no drug group on a two tailed Mann Whitney U-Test (Siegel, 1956). Neither the sodium seconal group nor the meprobamate group produced a mean RT significantly different from the placebo group. The data suggests that the mere administration of tablets produces a significant increase in RT and that the pharmacological properties themselves of sodium seconal and meprobamate at the dosages used do not produce a significant increase beyond this. These results are inconsistent with those of Kornetsky (1958) who found a significant increase in visual RT when the effects of meprobamate were compared to the effects of a placebo. He did, however, use a larger dose (1600 mg) than that used in the present study (800 mg) and psychophysiological measures have been shown to be sensitive to different dosages of meprobamate (Costello, 1962). Furthermore, Reitan (1957) did not find evidence of a significant change in visual RT after administration of 1600 mg of meprobamate.

The difference between the means for the D-amphetamine sulphate and no drug groups is not significant. Lehmann & Csank (1957) did find a significant shortening in RT after 26-28 mg of D-amphetamine sulphate. Neither 5 mg or 15 mg of D-amphetamine sulphate produced a RT significantly different from that of the placebo group in the Kornetsky study (1958).

Unfortunately the present results are difficult to interpret since the placebo group had a significantly longer mean RT than the no drug group and the



D-amphetamine sulphate group, suggesting the possibility that the lengthening of the RT produced by the psychological effects of the administration of tablets are offset by the pharmacological properties of D-amphetamine sulphate.

### DISCUSSION

The data do not confirm either of the predictions made since neither D-amphetamine sulphate or sodium secenal produce the paradoxical increase in RT with increase in stimulus intensity.

Venables & Tizard (1956b), in presenting a possible explanation for the disappearance of the paradoxical effect at the second session in their study, talk of 'emotional excitation' being present in the first session and not in the second session. They do not, however, distinguish between cortical and subcortical excitation.

On the other hand, Pavlov's discussion of paradoxical effects refers mainly to different levels of excitation at sub-cortical levels. The level of excitation of his dogs was determined mainly by numbers of hours of food deprivation (Pavlov, 1941). It is possible therefore that though production of high or low levels of excitation at a cortical level as was done in the present study by means of drugs may not result in paradoxical effects, manipulation of level of excitation at a sub-cortical level will result in such effects.

There is some evidence of a significant interaction between anxiety as measured by the Taylor Manifest Anxiety Scale (Taylor, 1953) and intensity of stimulus in determining RTs to an auditory stimulus (Castaneda, 1956). However, the results appear contrary to what one would predict on the above account of paradoxical effects since the significant interaction was based on a tendency for the speed of reaction of the high anxious group, in comparison to the non-anxious group, to be slower at the weak intensity but faster at the strong intensity. On the other hand, Venables & Tizard (1958) failed to find in schizophrenic Ss an increase in RT with an increase in auditory stimulus intensity. It is possible, then, that the results of the Castaneda study, apparently contradictory to prediction, are due to the sense modality used.

Further experiments in the present series will concentrate on visual RT, and level of excitation at a sub-cortical level will be manipulated by the use of Ss with extreme scores on anxiety scales, stress situations and tranquilizing drugs.

This work was supported by a grant from Wallace Laboratories, New Jersey. Thanks are due to Jim Stephen for technical assistance and to Gloria Glab who conducted the experiment.

### REFERENCES

- BERGER, G. O. (1886). Über den Einfluss der Reizstärke auf die Dauer einfacher psychischer Vorgänge mit besonderer Rücksicht auf Lichtreize. *Phil. Stud.* 3, 38-93.
- CASTANEDA, A. (1956). Reaction time and response amplitude as a function of anxiety and stimulus intensity. *J. abnorm. soc. Psychol.* 53, 225-228.
- CATTELL, J. M. (1886). The influence of the intensity of the stimulus on the length of the reaction time. *Brain*, 8, 512-515.
- COSTELLO, C. G. (1963). The Effects of Meprobamate on the After-Image. In *Experiments with Drugs* (ed. H. J. Eysenck). London: Pergamon Press.



- COSTELLO, C. G. & STEPHEN, J. The use of the tape recorder in psychological experimentation. *Amer. J. Psychol.* (To appear.)
- HULL, C. L. (1949). Stimulus intensity dynamism (V) and stimulus generalisation. *Psychol. Rev.* **56**, 67-76.
- KORNETSKY, C. (1958). Effects of meprobamate, phenobarbital and dextroamphetamine on reaction time and learning in man. *J. Pharmacol.* **123**, 216-219.
- LEHMANN, H. E. & CSANK, J. (1957). Differential screening of phrenotropic agents in man. *J. clin. exp. Psychopath.* **18** 222-235.
- PAVLOV, I. P. (1941). *Conditioned reflexes and psychiatry* (Trans. W. H. Gantt). London: Lawrence & Wishart.
- SIEGEL, S. (1956). *Nonparametric statistics for the behavioural sciences*. New York: McGraw-Hill.
- TAYLOR, J. (1953). A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.* **48** 285-290.
- VENABLES, P. H. & TIZARD, J. (1956a). The effect of stimulus light intensity on the reaction time of schizophrenics. *Brit. J. Psychol.* **47**, 144-145.
- VENABLES, P. H. & TIZARD, J. (1956b). Paradoxical effects in the reaction time of schizophrenics. *J. abnorm. soc. Psychol.* **53**, 220-224.
- VENABLES, P. H. & TIZARD, J. (1958). The effect of auditory stimulus intensity on the reaction time of schizophrenics. *J. ment Sci.* **104**, 1160-1164.

*Manuscript received 25 October 1962*

## Psychological Functions during Wave-Spike Discharge

BY BARBARA TIZARD AND J. H. MARGERISON

*Department of Experimental Neurology, Institute of Psychiatry, London, S.E.5*

The responses of six epileptics to a variety of psychological tests were studied during generalized WS discharge. All patients tended either to work more slowly or to cease work during WS lasting longer than 1 sec but some responses continued to be made even in discharges of long duration. The proportion of responses made during WS was shown to be related to the duration of WS, to individual characteristics of the patient and to the nature of the test. Modality of stimulus was not a significant variable. An attempt is made to explain the different numbers of responses occurring in different tests in terms of information theory.

### INTRODUCTION

A previous study (Tizard & Margerison, 1963) dealt with changes in the amount of Wave Spike (WS) discharge recorded from two epileptics during the performance of various tests and with changes in test performance during the occurrence of brief WS discharges. It was established that both patients worked more slowly and made more errors during WS than in its absence. Very brief bursts of 1 to 1½ sec duration, without overt behavioural accompaniments, were shown to be associated with significantly slowed response times. An attempt was made to relate the differences in the amount of WS recorded under different test situations to the degree of alertness of the subjects.

The present communication is concerned with an analysis of the responses during WS of larger numbers of patients in an increased variety of experimental situations. The results will be discussed in terms of information theory.

It is generally accepted that loss of consciousness is the usual clinical accompaniment of WS discharge of sufficiently long duration. Goldie & Green (1961) and others have described a typical 'trough-like' course of unconsciousness within a minor seizure. Responses are said to continue for the first few seconds of an attack and then totally cease until the last few seconds. It is, however, equally clear that cerebral functions are differentially impaired. The common clinical impression is that speech is more affected than is the carrying out of simple motor tasks, and this is supported by the work of Gastaut (1954) and of Schimanozo *et al.* (1953). However, the distinction between repetitive motor activity during seizures and automatic ictal behaviour is not always clear. In this investigation an attempt has been made to describe more precisely the disturbance of consciousness during WS. To this end patients have been examined with psychological tests of varying complexity, verbal and non-verbal, using three modalities of stimulus.

It was also hoped to explore the possibility that differences in test responses might be seen in attacks of varying duration. Little detailed attention has been paid in the past to the relationship between duration of WS discharge and behaviour during the discharge. It is, however, known that *petit mal* status may



present as intellectual retardation or as psychiatric disability, and in either case be without ictal features. Very brief paroxysmal discharges are also not usually associated with manifest clinical symptoms of epilepsy. In one of the few quantitative studies of this problem Schwab (1947) reports that response times to a simple auditory stimulus were delayed in attacks lasting 5 sec or less but that in those lasting more than 8 sec no response occurred and 'unconsciousness must be presumed'. In the present investigation patients with attacks of varying duration were included, in order to examine further the relationship between 'unconsciousness' and duration of WS discharge.

## METHOD

### 1. Subjects

These were selected from patients referred to the epilepsy service of the Maudsley Hospital in the course of a year. All patients were included provided that:

- (i) They suffered from very frequent minor seizures of presumed subcortical origin. They presented therefore with the clinical picture of 'true' *petit mal*. Patients with myoclonic *petit mal* were excluded as they were considered unlikely to be able to make test responses during seizures.
- (ii) Their EEGs showed bursts of bilaterally synchronous, generalized 3-4½ c/sec WS discharges. Minor asymmetries of background activity were tolerated but focal features were not.
- (iii) Each burst of WS discharge could be reasonably clearly demarcated with respect both to its beginning and to its end.
- (iv) They required admission to hospital for therapeutic reasons and were agreeable to taking part in the experiment.

Only six patients were found who satisfied these criteria. All suffered from occasional major as well as the much more frequent minor seizures and the attacks of *petit mal* with appropriate EEG accompaniments could be brought about in every case by hyperventilation. It is not suggested that the group is homogeneous except within the limits of the above criteria. Four patients were photosensitive whilst two were not, and there were individual differences in drug régimes.

The six subjects (*A, B, C, D, E* and *F*) fell into two categories with respect to the duration of bursts of WS. In three (*A, B* and *C*) the great majority of such bursts, hereafter called brief, lasted from ½ to 5 sec. Only when they lasted longer than about 3 sec were either clinical accompaniments seen or the subjects aware that an attack had taken place. The clinical features consisted of a hesitation in speech and a symmetrical fluttering of both eyelids. These WS discharges occurred very frequently, often two or three times a minute.

In the other three subjects (*D, E* and *F*) WS discharges, hereafter called long, were much less frequent, never more often than once every 3 min. In the case of *D* and *E* they usually lasted from 10-30 sec, in the case of *F*, from 30-120 sec. They were invariably associated with recognizable minor seizures in which patients would stop whatever they were doing, fail to answer questions, and assume a blank expression. Involuntary movements of the limbs sometimes occurred.

Whilst none of the 'brief' group had bursts of WS lasting so long as 10 sec and none of the 'long' group had bursts lasting less than 5 sec, all subjects had occasional discharges lasting from 6 to 8 sec.

Further particulars of the patients are supplied in Table I.

### 2. Procedure

EEG (five channels), ECG, respiration, test stimuli and responses were all recorded with a Mark III, eight-channel, Edison-Swan EEG machine (see Fig. 1). Testing was performed under standard conditions, at the same time each day, for period of 30 to 45 min, with the subject seated comfortably at a table throughout. At least ten records, each containing spontaneous WS discharges, were obtained from every patient.



Table 1. *Subjects of the investigation*

	Age	Sex	Full Scale W.B.I.Q.
A	23	F	86
B	62	F	64
C	10	F	104
D	21	M	88
E	15	M	110
F	14	M	105

### 3. Psychological Tests

All these were simple, repetitive and continuous. Ten minutes was usually spent on each test on any one occasion. A short practice period preceded testing each day. Only two subjects carried out all the tests described below because the battery of experiments was gradually increased as data from successive patients were analysed. This procedure was considered to be justifiable, indeed necessary, as suitable subjects were difficult to find and it was felt that the opportunity to obtain as much useful information as possible should not be lost.

(a) *Tape Test.* All subjects did this. A recorded series of numbers, from one to nine, randomly presented, was read out at the rate of 2 per sec. The task was to press a rubber bulb as soon as possible, whenever the number 'six' was heard. Errors of omission and of

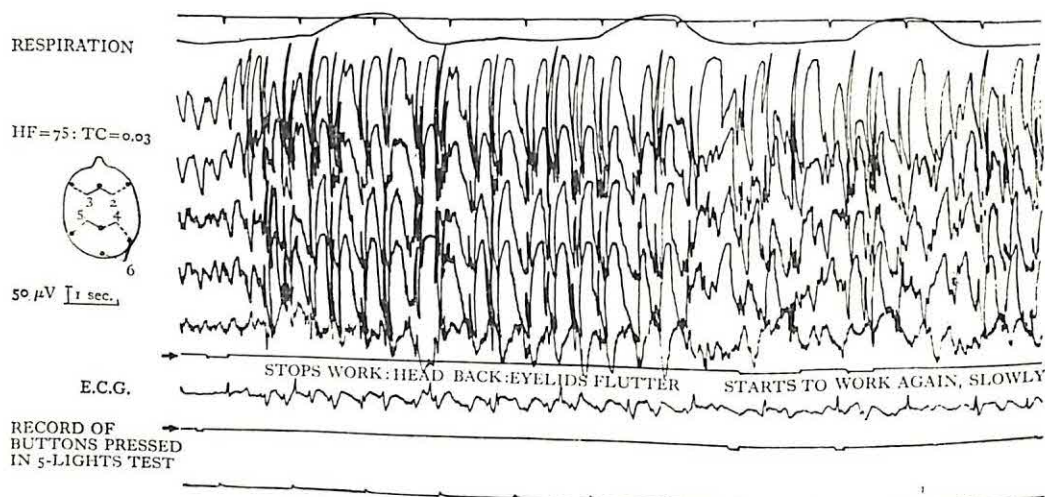


Fig. 1. Tracing from 'D' at start of minor seizure.

commission were scored. *D* and *E* also did a slower version of the test in which the numbers were read out at the rate of 1 a sec. It was found that during an attack subjects *D* and *E* therefore used for these patients, namely, lifting the forearm, making a pencil stroke, saying 'Bone'.

(b) *The 5-Lights Test.* Five subjects, *A*, *B*, *C*, *D* and *E*, did this test. Five small light bulbs were arranged in a semicircle on a board. When a centre button was pressed one or other of the bulbs lit up. The light was put out by pressing a button adjacent to it. Pressing the centre button again caused another bulb to light up. The subject was instructed in putting on and off the bulbs and asked to repeat the sequence as quickly as possible until told to stop. The order in which the bulbs lit up was randomized. Two sets of scores were used:

- (1) Response times, i.e. the intervals between successive pressings of the centre button, measured in mm (1 sec = 3 cm).

- (2) Errors, i.e. pressing the wrong peripheral button.

One subject, *D*, also did a 'paced' version of the test. In this version the peripheral lights automatically came on and were extinguished at the rate of 30 a min, whether or not the subject pressed the buttons in time.

- (c) *Simple and choice reaction tests.* Three subjects did these tests: *D*, *E* and *F*.

(a) *Visual Reaction Tests.* A small wooden box in which was set a ground glass screen, 9 in.  $\times$  7 in., was placed before the subject. The visual stimuli were coloured lights which illuminated the whole screen. In the simple test only red lights were used and the subject was asked to squeeze a rubber bulb as soon as he saw the light. In the choice test red and blue lights were presented in random order and the subject was asked to respond only to the red light. The testing room was lit by electric light and maintained at approximately constant illumination throughout the investigation.

(b) *Auditory Reaction Tests.* Tones, subjectively assessed as 'loud' and 'rather soft', were relayed by a loudspeaker placed before the subject. In the simple test only the loud noise was used and the subject was asked to respond as soon as he heard the noise. In the choice test loud and soft noises were randomly presented, and the subject was asked to respond only to the loud noise. *E* also did a version of this test in which two different tones were presented in random order.

(c) *Tactile Reaction Test.* A tactile stimulator was strapped on to the subject's palm and he was asked to respond with the other hand as soon as he felt the tickling sensation. Only the simple version of this test was used.

The timing of all the stimuli was controlled electronically. All stimuli lasted 0.5 sec and they were presented at intervals of 2-3 sec. Errors of omission and commission were scored.

- (d) *Verbal commands and questions.* These were used in the investigation of one patient, *E*, and will be described under 'Results'.

#### 4. Reliability of WS measurement

Test responses were correlated with WS discharge, rather than with the clinical attack. Criteria for the beginning and end of discharges were agreed upon initially. Since only patients with reasonably clearly demarcated WS bursts were selected, it was then possible to obtain a high measure of agreement between two observers on the duration of WS bursts. In the case of one patient, *B*, the reliability of measuring individual discharges was assessed. A correlation of +0.94 was found between two independent estimates of the length of ten discharges. It should, however, be pointed out that the end of a WS burst is less abrupt than its beginning, and the end points chosen were therefore to some extent arbitrary.

## RESULTS

### *Tape Test*

When WS lasted for  $1\frac{1}{2}$  sec or more all subjects tended to omit responses.

(a) *Long discharge group.* *D*, *E* and *F* made no responses at all during WS in the quick or the slow version of the test unless the '6' was spoken in the final second of WS. This obtained irrespective of the type of response.

(b) *Brief discharge group.* *A*, *B* and *C* sometimes responded to '6s', but in WS lasting  $1\frac{1}{2}$  sec or longer made significantly more omission errors than during the rest of the test. The records were examined to see whether omission errors were made less frequently in the first and last second of WS than in the central part. (Only WS lasting 3 sec or longer was considered.) There was no difference between the frequency of omission errors in the first, second and the central part of the discharge, but omissions were rare if the '6' was spoken in the last second of WS. Very few commission errors were made by any subject and there was no tendency for these to increase during WS.

(c) *Individual differences.* The performance of *A* was much less affected by WS



than that of *B* and *C*; *A* responded to 80 per cent of the '6s' that occurred during WS, *B* and *C* to only 22 and 20 per cent respectively. This difference cannot be attributed to differences in the duration of WS, since it remains almost the same if only WS lasting 3-4½ sec is considered.

Table 2. *Errors and WS*

	Tape Test		$\chi^2$
	Omissions during WS	Omissions during periods without WS	
<i>A</i>	35 (1,041 sec)	200 (10,959 sec)	11.3†
<i>B</i>	450 (2,180 ,, )	195 ( 8,620 ,, )	1169.0‡
<i>C</i>	39 ( 170 ,, )	893 ( 7,630 ,, )	9.6†

	Five Lights Test		$\chi^2$
	Errors during WS	Errors during periods without WS	
<i>B</i>	97 (WS = 716 sec)	230 (10,084 sec)	77.0‡
<i>D</i>	14 (WS = 236 ,, )	14 ( 2,164 ,, )	51.0‡

\*  $p < 0.05$ ; †  $p < 0.01$ ; ‡  $p < 0.0001$

## 2. 5-Lights Test

When WS lasted for 1 sec or more all subjects tended to work more slowly or to stop working.

(a) *Long discharges group.* *D* and *E* both tended to make errors and stop work for varying periods of time. *D* stopped work altogether in WS of 7-12 sec ( $N = 5$ ), from the first WS until the discharge was over. However, when WS lasted from 13-44 sec ( $N = 12$ ) after an initial pause of 6-8 sec he started to work again although slowly and making many more errors than usual. Similar response patterns were observed in the 'paced' version of the test. *E*, on the other hand, irrespective of the duration of WS (which was from 6-27 sec), always continued to work for the first 1-3 sec of WS, then made errors and finally stopped work altogether. In six out of nine discharges he started to work again in the final 1-3 sec of WS.

(b) *Brief discharges group.* All subjects worked significantly more slowly during WS, but only one, *B*, made significantly more errors during WS than during the rest of the test. Table 3 shows that with *A* response times were slowed during WS lasting only ½ sec. In all cases WS of only 1-1½ sec was accompanied by a highly significant slowing of response time.

## 3. Simple Reaction Tests

These were done only by the patients with long WS discharges. They responded to some, but not all of the stimuli presented. Modality of stimulus had no significant effect on the proportion of responses omitted by any subject in both simple and choice reaction tests. There were marked individual differences in responsiveness. *F* had only one WS discharge, of 88 sec, during these tests. He responded to only one of the stimuli presented, the second. *D* made no response to the first 2-4 stimuli occurring in WS (duration 6-31 sec) but thereafter responded intermittently in the longer discharges. *E* responded to many more of the stimuli



Table 3. *Response time and WS*

	During WS $\frac{1}{2}$ -1 sec	During periods without WS	During WS 1-1 $\frac{1}{2}$ sec	During WS 3+ sec
A				
Mean response time (in mm)	43.33	38.30	48.28	79.47
S.D.	8.83	9.60	17.73	40.00
		$t = 3.29\ddagger$	$t = 5.03\ddagger$	$t = 5.61\ddagger$
B				
Mean response time (in mm)	—	37.46	55.72	61.30
S.D.	—	4.60	14.35	32.73
			$t = 16.02\ddagger$	$t = 1.88$ N.S.
C				
Mean response time (in mm)	38.2	44.12	63.62	63.24
S.D.	16.82	13.50	32.46	32.30
		N.S.	$t = 3.48\ddagger$	N.S.

(All  $t$ -tests made on logarithmically transformed data)

presented during WS (duration 8-31 sec). He always responded to the first stimulus generally missed the second, but thereafter responded to all stimuli. No changes in the form of the EEG discharge could be seen to correspond to these fluctuations in responsiveness.

Table 4. *Failure to respond to Reaction Tests during WS*

	Choice Tests		Simple Tests		$\chi^2$
	No. of bursts without response	No. of bursts with response	No. of bursts without response	No. of bursts with response	
D					
	14	5	14	18	4.40*
E	8	9	2	17	4.05*

#### 4. Choice Reaction Tests

These also were done only by the patients with long WS discharges. *D* and *E* responded to some of the stimuli correctly but there was a significantly larger proportion of discharges in which no response at all was made (except to stimuli in the last second) than during the simple reaction tests. Incorrect responses (i.e. to the 'wrong' stimulus) were almost never made; the tendency was to omit responses during WS rather than to respond inappropriately. The individual differences in responsiveness that had been observed in the Simple Reaction Tests were observed again. *D* made no response in WS of 6-15 sec. In longer discharges of up to 30 sec after an initial period of unresponsiveness intermittent correct responses were usually made. *E* made no response in 8 out of 17 discharges but in the remainder generally responded correctly to all but the first stimulus. The duration of WS (6-22 sec) did not appear to be a factor affecting his responsiveness, nor did the substitution of two different tones for two different intensities of sound affect the number of his responses. The proportion of WS discharges during which no response was made to any stimuli was significantly larger in the case of *D* than *E* even if only WS of 10-20 sec is considered. *F* had only one WS discharge lasting

52 sec during this test, in the course of which he made only one correct and one incorrect response, both in the middle of the attack.

Table 5. *Individual differences in failure to respond to Reaction Tests during WS*

	No. of attacks without response	No. of attacks with response	$\chi^2$
<i>D</i>	28	23	
<i>E</i>	10	26	6.8*

#### 5. *Speech during WS discharge*

In all our patients spontaneous speech invariably stopped during clinical attacks. We assumed that speech as such was impossible for them during WS. However, whilst investigating our last two patients, *D* and *E*, it was discovered that both could continue to make a verbal response (any pre-arranged word, e.g. 'bone') to the simple and choice reaction tests during WS. In the case of *E* imitative speech was also possible. This patient if told, after WS had started, to repeat a phrase or list of four or five words would do so after a short delay. However, a reply to the simplest question, e.g. 'What is your name?' was never given. Similarly he could repeat five digits correctly during the course of WS but was unable to answer such questions as 'What number comes after 5?' or ' $2 + 2 = ?$ '

#### 6. *Memory processes during WS*

All but one of our patients had amnesia for events occurring during clinical attacks. *E*, however, claimed to be aware of what was going on around him and subsequently to remember it. He said that he could hear the '6s' spoken during the tape test, although he could not respond to them, and he was usually able to repeat the questions spoken to him during WS even though he had not answered them. It was decided to test his memory for non-verbal events during WS. After WS had begun sequences of red and blue lights and loud and soft noises were presented and after WS had finished he was asked to recall the order of presentation. This experiment was repeated several times. He was able to recall a series of up to six stimuli correctly, e.g. one red, three blue, two red, although during WS he had not responded to them correctly. An investigation of his memory for verbal stimuli during WS was next undertaken. It was discovered that only brief familiar remarks or questions were correctly recalled; if connected prose, e.g. a passage from a children's book, was read aloud during an attack only the last few words were recalled. More significantly, perhaps, unfamiliar questions such as 'Do houses float?', 'Do books eat?', were recalled only partially and incorrectly.

### DISCUSSION

These findings suggest that the behavioural correlates of WS are quite inadequately described as 'loss of consciousness during WS of sufficiently long duration'. On the one hand, brief discharges, during which no seizure is reported by the patient or noted by a skilled observer, may, in fact, be accompanied by a significant slowing in response time and a tendency either to omit responses or to make errors. It has been suggested elsewhere (Tizard & Margerison, 1963) that these behavioural



correlates should, in the presence of WS discharge, constitute acceptable evidence of a seizure even though no change be observed in the patient by a competent clinician. The possible handicap of these brief 'subclinical' seizures should be considered in assessing the achievements and psychological test results of epileptics. On the other hand, during long discharges unexpectedly selective behaviour may be possible. Despite their apparent 'absence' and subsequent amnesia, patients may be able to discriminate between two lights of different colour and between loud and soft sounds.

However, although the patients with 'long' discharges continued to make some responses during WS, these were fewer in number than those made by patients with 'brief' discharges. Moreover, the 'long' discharges were accompanied by a different level and pattern of responsiveness from their very beginning and were in this sense different in kind, and not merely in duration, from the 'brief' discharges. This is illustrated by the responses during WS bursts of 6-8 sec which were seen occasionally in the EEGs of each of the patients studied. The group with 'brief' WS responded to some of the '6s' and worked on slowly at the 5-Lights Test, but the long duration WS group did not respond to any of the '6s' and ceased work on the 5-Lights Test. Duration of WS is, however, not the only factor affecting responsiveness during attacks. If only WS of 3 to  $4\frac{1}{2}$  seconds and 10 to 20 seconds is considered, there were marked individual differences in the proportion of responses missed. Our data suggest, therefore, that the response level during WS is not a direct function of the duration of WS, but that both are determined by some other mechanism.

The distribution of responses during WS did not follow the 'trough like' course often described. Among those patients with brief WS, omissions were neither more nor less frequent at the beginning than in the middle of the discharge but occurred very rarely if the '6' was spoken in the final second. Two of the three patients with long discharges (*D* and *F*) ceased responding with the first WS. *D* resumed work after varying periods of time, the length of which appeared to depend in part upon the nature of the test. Often there would be further brief interruptions of activity during the course of WS. The remaining patient (*E*) continued to respond for the first few seconds of WS, stopped briefly and then resumed without further interruptions provided that the test was relatively easy. When the tasks were more difficult he either stopped work completely or alternatively began to respond again a few seconds after the beginning of WS. Each patient had a 'style' of response during WS, and the number and pattern of his responses during different tests were very consistent.

Individual differences, in fact, are as striking in this series as are group characteristics. It was stated earlier that although clinical and electrical criteria were strictly applied the group was not considered to be homogenous except within the limits of these criteria. In addition to the differences already described between patients with respect to age, I.Q., duration and frequency of WS, behaviour during WS, response to photic stimulation and to drugs, it should be pointed out that the qualitative appearance of the WS discharge was recognizably different from subject to subject. Our series is probably quite representative of *petit mal* epileptics, i.e.



those with minor seizures of presumed subcortical origin, and it may be worth emphasizing that these patients form in many respects a very heterogeneous collection.

It was noteworthy that, although changes in test responses always occurred during the course of WS discharges, no corresponding alterations in the EEG record could be detected by visual inspection. There were often minor alterations in recurrence rate of the individual WS complex, variations in amplitude, and the phase relations of fast and slow components sometimes shifted, but these changes did not appear to correspond in any consistent way to alterations in behaviour. This negative finding may, of course, simply reflect the crudeness with which the electrical changes were assessed.

A consistent finding was that during WS response failures occurred more often in some tests than in others. Thus delayed reactions or omissions errors occurred most often during the Tape Test, less often during the Choice Reaction Test, less often still during the Simple Reaction Test, and least of all during the 5-Lights Test. Spontaneous speech was always arrested, but some verbal responses were made. This pattern of response failure could not be explained in terms of the modality of the stimulus, although it is tempting to speculate that this might prove otherwise with minor focal cortical seizures. However, the data become meaningful if interpreted in the light of certain concepts taken from information theory.

The value of one such concept, capacity, in considering brain function has been emphasized by Broadbent (1958). The 'capacity' of the brain is the limit to the quantity of information which it will handle in a given time. Information varies directly with the rate at which the stimulus is presented, and inversely with the probability of the stimulus. The hypothesis put forward here is that during WS there is a temporary reduction in the 'capacity' of the brain.

If the reduction in capacity is small, only the rate of response may be affected. If the reduction in capacity is greater only messages of low information value can be transmitted. For this reason responses are made more often in simple than in choice reaction tests for, although stimuli and responses are identical, more information is presented with the stimuli in the choice reaction tests. Fewer responses still are made in the Tape Test since each stimulus comes from an array of nine digits.

In the case of speech, not only does each word carry a large amount of information, but all the information in a sequence of words must be handled together. It is probably for this reason, rather than because of any specific qualities of speech, that answering questions is impossible during WS. Two of our patients, in fact, used low information content speech during WS. Patient *E* could even repeat 5 digits correctly during WS yet often failed to respond in a choice reaction test. Although the former task appears superficially to be more 'difficult' it involves only short-term storage and subsequent reproduction of a series of stimuli whilst in the choice reaction test the stimuli involve uncertainty of response.

Patient *E* also said that he could perceive during attacks and could remember afterwards what had happened. This statement appeared to present a considerable problem to our theory, because if perception and memory remain intact during

WS then only a transient arrest of voluntary motor activity need be postulated. However, investigation showed that stimuli of relatively high information value (unfamiliar phrases and continuous prose) were not, in fact, remembered after an attack. Moreover, there was no arrest of motor activity or even speech as such but only of responses to stimuli of high information content. The fact remains that for this particular patient, at least, response mechanisms appeared to be particularly affected by WS discharge. Memory functions, although less efficient than usual, were better preserved, for a sequence of lights could be recalled although no response had been made to them.

It is not suggested that the simple theory which we have put forward offers an adequate explanation for the complex changes in psychological functions which can be clearly demonstrated during the WS discharges. It none the less enables us to give some rational explanation of the results which we have obtained and has the merit that it can lead to further experiments.

This work was carried out whilst the first author was in receipt of a grant from the Medical Research Council.

A grant for apparatus from the British Epilepsy Association is gratefully acknowledged. We wish to thank Professor G. D. Dawson and Dr D. A. Pond for their encouragement and help and Dr D. W. Liddell of the Maudsley Hospital and Dr J. Bowe of Lingfield Hospital School for permitting us to study their patients. We are also very much indebted to the staff of the EEG Department and the nursing staff of the Maudsley Hospital for their assistance and to Mr P. St John-Loe for making and maintaining apparatus.

#### REFERENCES

- BROADBENT, D. E. (1958). *Perception and Communication*. Oxford: Pergamon Press.
- GASTAUT, H. (1954). In *Brain Mechanisms and Consciousness* (ed. Delafresnaye, J. F.). Oxford: Blackwell.
- GOLDIE, L. & GREEN, J. M. (1961). Spike and wave discharges and alterations of conscious awareness. *Nature, Lond.* **191**, 200-201.
- SCHIMAZONO, Y., HIRAU, T., OKUMA, T., FUKUDA, T., & YAMAMASU, E. (1953). Disturbance of consciousness in *petit mal* epilepsy. *Epilepsia*, **2**, 49.
- SCHWAB, R. S. (1947). Reaction time in *petit mal* epilepsy. *Ass. Res. Nerv. Ment. Dis.* **26**, 339-341.
- TIZARD, B. & MARGERISON, J. H. (1963). The relationship between generalised paroxysmal EEG discharges and various test situations in two epileptic patients. *J. Neurol. Neurosurg. Psychiat.* **26**, 308-313.

*Manuscript received 20 November 1962*



## Cross-masking in Neurotic Patients\*

By J. G. INGHAM

*Medical Research Council, Social Psychiatry Research Unit,  
Llandough Hospital, Penarth, Glam.*

### EARLIER WORK

#### (a) *Results*

Cross-masking is the elevation in monaural threshold which occurs in the presence of a tone of different frequency in the opposite ear (Ingham, 1957; Chocholle, 1960). In certain respects, it resembles phenomena of 'external inhibition' reported by Pavlov. Most of the recent work on the predominance of inhibition in hysterics compared with dysthymics has involved phenomena of 'internal inhibition'. There is no reason to suppose that the two kinds of events involve similar processes. Pavlov (1932), in his later writings, seems to have had 'external inhibition' in mind when he referred to the predominance of inhibitory processes in hysterics. The aim of the present work was to see whether there was any evidence of this for cross-masking.

A first inquiry (1958), using the procedures of the earlier investigation (1957), showed a significantly greater masking effect in hysterics than dysthymics. The frequency of the masking tone was 400 c/s, 30 dB above threshold, and the frequency of the test tone was 1000 c/s throughout the session.

Two later investigations were intended to see whether the observation could be extended to different methods of measuring the masking effect and to different test frequencies. Although the differences observed were in the predicted direction, they were not significant. For one group of patients, the procedure was thought to be substantially the same as that used previously, yet dysthymics and hysterics did not differ significantly and both groups showed significantly less masking than before ( $p < 0.02$  in both cases). This was now made the starting point of a further investigation designed to discover the reason for the discrepancy.

#### (b) *Changes which could account for the discrepancy*

A re-examination of the experimental conditions suggested that two changes, introduced after the 1958 investigation, might have modified the masking effects differentially. One was the introduction of an oscillator for generating the masking tone, to replace the tape recorder used previously. The oscillator produced a sine wave with little distortion, whereas the tape-recorder gave a wave-form which was distorted and varied slightly in amplitude and frequency. There was evidence from another experiment that the tape recorder caused a greater masking effect than the oscillator in the same individual.

The second of the changes concerned the procedure for threshold measurements. A method of limits was used in which the intensity of the test tone was increased in steps of 1 dB until the subject reported its presence. In the 1958 investigation the increments continued, if no response occurred, only until the tone was 10 dB louder than the previously recorded threshold. Quite often, during determinations of masked thresholds, there was no response. To overcome the difficulty of having no accurate measurements in these instances, the procedure was altered to allow increments in the test tone until it was 20 dB louder than the previously recorded threshold. With the new procedure, in the later investigation, complete failures to hear the test tone were rare. When the earlier results were scrutinized, it became apparent that the difference between hysterics and dysthymics might have existed

\* A detailed account of all these investigations is available from the author on request.



only in those patients who failed to hear one or more test tones. In such instances the interval between two successive responses was inevitably lengthened. It seemed possible that it was the influence of this upon subsequent thresholds that differed from hysterics to dysthymics. By changing the procedure, an essential condition for the occurrence of the difference might have been removed.

### PRESENT INVESTIGATION

#### (a) Method

It was now necessary to find out whether the change in masking stimulus had reduced the difference between the groups or whether it was the change in procedure or a combination of both. Three groups of patients were studied, using a different combination of masking tone and procedure for each group. These were:

- (a) Old procedure and oscillator.
- (b) New procedure and tape recorder.
- (c) Old procedure and tape recorder.

The remaining combination of new procedure with oscillator was omitted for economy.

A standard diagnostic interview was conducted with each patient,\* who was then labelled according to criteria stated by Shagass & Naiman (1956). For the first stage of the analysis, two broad categories were used. The hysterics included both conversion hysterics and hysterical personalities. The dysthymics included anxiety neurotics, neurotic depressives, obsessive compulsives and phobics. Ages ranged from 17-65 years. Patients within each broad category were allotted at random to the three experimental groups (a), (b) and (c).

#### (b) Results

Results are summarized in Table 1. The difference between the mean masking effects of the hysteric and dysthymic groups was of the same order for female patients as the 1958 results. This was true of all three experimental conditions.

Table 1. *Average masking effect in decibels (M) and number in group (N)*

	(a)				(b)				(c)			
	Male		Female		Male		Female		Male		Female	
	N	M	N	M	N	M	N	M	N	M	N	M
Dysthymics	18	4.3	13	5.5	18	4.1	12	5.2	20	4.3	14	7.0
Hysterics	9	3.8	8	8.3	9	5.3	7	8.9	9	3.7	9	9.7

Correcting for age by analysis of covariance, the F-ratio for diagnosis in the female group was significant at the 5 per cent level. For male patients, however, there was no evidence of such a difference between groups. There appears to be a difference between hysteric and dysthymic women under all the experimental conditions, but this does not hold for men. When the results of previous investigations were re-examined it was found that they did not contradict such a conclusion. Only the 1958 sample contained sufficient female hysterics to provide a reliable estimate of mean masking effect in the female hysteric group. The figure obtained was 9.1 dB, about the same as in the present inquiry. The mean masking effect in male hysterics was also high (7.5 dB) but there were only three people in this group and their high scores are easily attributable to sampling errors. The possibility of an interaction between sex and diagnosis was overlooked because of inadequate numbers in one group, and this was also true of later investigations. There was no significant variation between experimental conditions for either sex.

\* Details of the interview are available on request.

One possible explanation of the difference between the sexes is that different criteria of diagnosis were used. To examine this possibility the results were re-analysed without using the diagnoses, but only the primary information from which the diagnoses were derived. A separate analysis of covariance was done for the nine items of information from the interview records which had contributed significantly to the diagnosis in either sex. If the absence of an association between diagnosis and masking in the men was only due to different diagnostic criteria having been applied, then the primary information should show the same relationships with masking in both sexes. In the female group, clear evidence was, in fact, found of an association between masking and several items of the primary interview information. Three of these were statistically significant. Depression was the clearest example (see Table 2). Analysis of covariance using a classification based

Table 2. *Average masking effect in decibels (M) and number in group (N)*

	(a)				(b)				(c)			
	Male		Female		Male		Female		Male		Female	
	N	M	N	M	N	M	N	M	N	M	N	M
Depressed	12	3.4	14	5.3	15	4.3	11	4.8	13	4.6	15	5.9
Not depressed	15	4.7	7	8.9	12	4.8	8	8.9	16	3.7	8	12.2

upon this symptom alone, instead of diagnosis, showed a highly significant difference in masking effect between depressed and not depressed women ( $p < 0.001$ ). In the male group, however, there was no evidence of an association between masking and any symptom or characteristic contributing to the diagnosis. The means shown in Table 2 for depression in the male group are typical. It is evident, therefore, that the differences in diagnostic criteria which do exist cannot account for the fact that different results were obtained from the male and female groups.

## DISCUSSION

Perley & Guze (1962) have suggested that hysteria, properly defined, is a condition which only occurs in women. Their definition requires that to qualify for a diagnosis of hysteria a patient must have, among other things, symptoms distributed among at least nine out of ten specified groups. One of the groups includes only menstrual disturbances of various kinds. Another comprises 'sexual indifference, sexual frigidity, dyspareunia, other sexual difficulties or vomiting for all nine months of pregnancy'. If, by definition, hysterics have at least one of these symptoms, it is clear that male hysterics must be very infrequent.

Information was not available for all the symptoms listed by Perley and Guze, but all the women in the present work were asked about menstrual disturbances and sexual difficulties. There was a slight indication among dysthymic women that the presence of these symptoms was associated with masking but this was in the opposite direction for hysterics. Probably the most crucial observation was that a group of nine female hysterics who had neither menstrual nor sexual symptoms, and therefore would not be considered hysterics on the Perley and Guze criteria,



still had a mean masking effect of 11.7 dB. Similarly, there was an average masking effect of 10.7 dB for nine women who were not depressed and had neither menstrual nor sexual symptoms.

Comparative information from a random sample of normal individuals is not yet available, but a small group of eight female hospital employees had a mean masking effect of 5.4 dB. This suggests that it is the hysteric women, and those who are neurotic but do not complain of depression, who differ from the normals and dysthymics but a larger and better sample may show otherwise. There is no evidence that masking is associated with extraversion as assessed by the MPI (Eysenck, 1959). Correlations varied from  $-0.31$  to  $+0.37$  in groups of about 20 individuals.

The results for women give some support to the Pavlovian hypothesis that inhibitory processes predominate in hysterics. There is evidence that hysteric women behave, in the masking situation, according to prediction, but a satisfactory theory will have to account for the fact that the men do not.

I am indebted to Dr J. P. Spillane and to the patients and staff of Whitchurch Hospital, Cardiff for their co-operation. This work was started while the writer was on the staff of the M.R.C. Neuropsychiatric Research Unit. I wish to thank the Director, Dr D. Richter, for allowing me to use the apparatus until the investigation was completed. Standard Telephones and Cables, Limited, Newport, kindly provided time on their computer.

#### REFERENCES

- CHOCHOLLE, R. (1960). Le seuil auditif monaural en presence d'un son de fréquence différente sur l'oreille opposée. *J. Physiol., Paris*, **52**, 51-52.
- EYSENCK, H. J. (1959). *Manual of the Maudsley Personality Inventory*. London: Univ. London Press.
- INGHAM, J. G. (1957). The effect upon monaural sensitivity of continuous stimulation of the opposite ear. *Quart. J. exp. Psychol.* **9**, 52-60.
- INGHAM, J. G. (1958). Auditory masking in neurotics. *Bull. Brit. Psychol. Soc.* **35**, A 18-19.
- PAVLOV, I. P. (1932). Essay on the physiological concept of the symptomatology of hysteria. In *Selected Works*, 1955. Moscow: Foreign Languages Publishing House.
- PERLEY, M. J. & GUZE, S. B. (1962). Hysteria—the stability and usefulness of clinical criteria. *New Engl. J. Med.* **266**, 421-426.
- SHAGASS, C. & NAIMAN, J. (1956). The sedation threshold as an objective index of manifest anxiety in psychoneurosis. *J. psychosom. Res.* **1**, 49-57.

*Manuscript received 30 November 1962*

## The Effect of Visual Distraction on Perception in Subjects of Subnormal Intelligence

By ROY I. BROWN

*Psychology Department, Hortham-Brentry Hospital Group, Bristol*

Distractibility is a term often used in clinical psychology though its meaning is ill-defined and little is known about the conditions under which it occurs.

A visual distraction task was given to 40 subjects of below average intelligence. A number of hypotheses related to brain damage, early environment and intelligence were investigated. Brain damage and low intelligence were not found to be associated with a high incidence of distractibility though early, and possibly long, institutionalization appeared to be important variables. The results are discussed in relation to recent work on deprivation.

### INTRODUCTION

Many of the terms in clinical psychology are often ill-defined or are so general in meaning as to be of little value. Distractibility seems to be one such term. A patient is very often described as distractible yet whether he is distracted under all conditions or at all times is not known. It has often been said that brain damaged subjects are distractible, for example, Strauss & Lehtinen (1960) indicated that brain damaged children show disturbance in learning and emotional behaviour as a result of their organic injury, and that part of the disturbance is observed in their degree of distractibility and their inability to select relevant from irrelevant stimuli. They indicated that brain damaged children are distractible in auditory, visual and tactile spheres. The work of these two authors has been severely criticized in recent papers, for example Pond (1961), in discussing epileptics, suggested that it is only those subjects from disturbed environments who have marked behaviour disorders. Woodward (1960) has shown that amongst severely subnormals distress behaviour (including fluctuating attention) tends to occur most frequently amongst those who come from poor or bad social backgrounds. These two papers suggest that behaviour disturbances in brain damaged subjects may not be determined by organicity alone, but are also affected by the type of background from which the subject comes. In addition, Brown & Clarke (1963) noted that in a group of non-Mongol imbeciles who were subjected to auditory distraction, the degree of distractibility varied considerably, some of the subjects being no more distractible than others of much higher intelligence. The authors suggested that this might be due to a greater heterogeneity of aetiology in the imbecile group. Thus it appears that a number of variables may be relevant in accounting for distractibility. It would seem appropriate to investigate such aspects as intelligence, aetiology and social background and also the subject's previous experience. In the distraction situation the length of task, the amount of learning required during the task, and also the modality in which a task is performed, should come under observation.



Further, the type of distraction used in relation to the task itself may be of considerable importance since it is possible that cross modality interference may be different from that within any one modality. It would seem worthwhile investigating whether a person distractible under visual tasks is equally distractible under auditory tasks.

The present experiment is concerned with visual distraction in subjects of below average intelligence. A number of hypotheses were put forward:

1. Not all brain damaged subjects would be distractible.
2. In view of the work of Woodward and Pond, early environment would influence the degree of distractibility. Two aspects of environment were recorded:
  - (a) The age at which institutionalization took place and its duration.
  - (b) The incidence of very bad and less adverse home backgrounds according to Clarkes' (1954) criteria.
3. Those subjects who showed a Performance I.Q. below Verbal would be more visually distractible than those who had a Verbal I.Q. lower than Performance.

#### METHOD AND PROCEDURE

Each subject was presented with 12 cards one at a time. Figures of varying complexity were shown on each card, e.g. circle, diamond, square with cross (see Fig. 1). The number of figures on each card ranged from one to four. The cards were presented on either a white

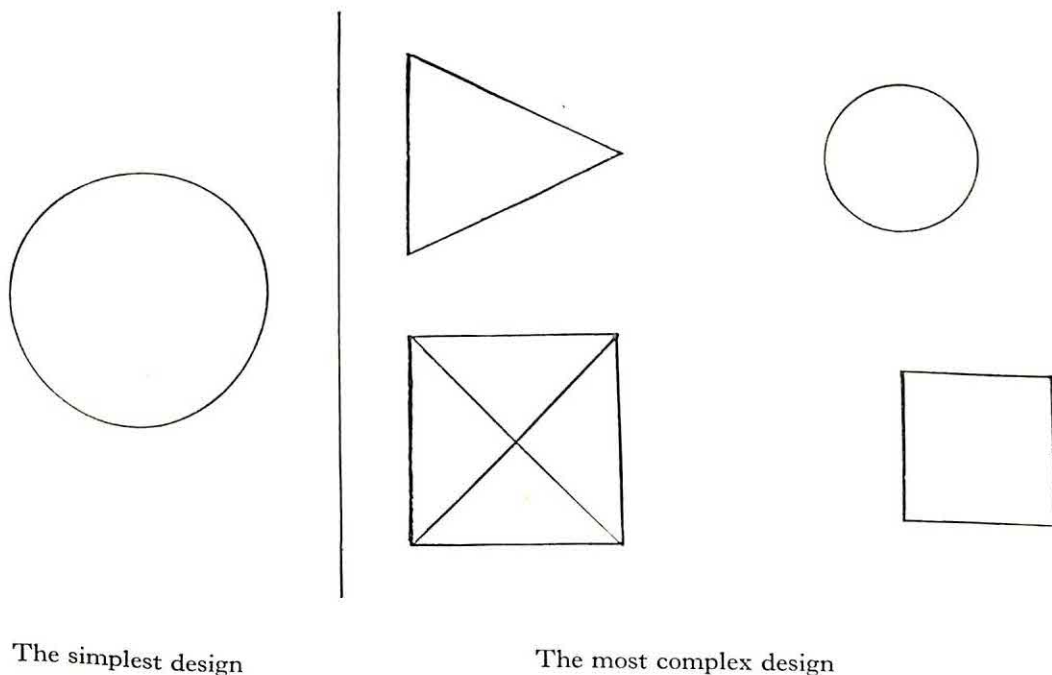


Fig. 1. Examples of stimulus cards.

background or on a surround containing 37 geometric figures. In the latter case the stimulus card was clearly delineated from the distracting surround. All 12 cards were given twice, once with and once without the distraction surround, and the distraction and non-distraction series were balanced for position and practice effects. This resulted in four groups with



different orders of presentation; each subject was allocated at random to one of these groups. Each card was presented for 3 sec after which the subject was immediately required to draw the stimulus pattern. The reproductions were scored as follows:

- |  |           |
|--|-----------|
| 1. Each figure up to the number on the stimulus card   | Score 2   |
| Additional figures   | Score - 2 |
| 2. Each correct figure   | Score 2   |
| A figure was counted correct provided it approximated in structure to the stimulus figure. For example, any three sided figure was accepted for an equilateral triangle.   |           |
| 3. Each correct figure in the right position   | Score 2   |
| A figure was considered in the right position provided part of the figure overlapped the stimulus figure when the response card was superimposed on the stimulus card.   |           |
| 4. The correct figure rotated  | Score - 1 |
| A figure was regarded as rotated when the majority of lines in a figure varied in direction from the stimulus figure. In a few borderline cases distraction and non-distraction figures were compared and a rotation score given to the one showing the greatest variation from the stimulus figure. |           |

The total score of the non-distraction trials was subtracted from the distraction totals—thus negative scores indicate distractibility. Since the weighting given to the four aspects of drawing was arbitrary, it was decided only to refer to scores as negative or positive, thus minimizing the possibly spurious effect given by some large positive and negative scores.

### *Subjects*

The subjects were below average intelligence and hospitalized at the time of research. As can be seen in the tables, the groups compared appear reasonably well balanced in terms of intelligence and age. In addition, an analysis of the effects of these two variables suggested that they played a minimal role in distractibility in the experiment (see later).

## RESULTS

Twenty-one subjects gained negative and 19 gained positive scores. There was one zero result. Thus 52.5 per cent of the subjects decreased their scores and it would appear that there is no overall factor of distractibility under the conditions of the experiment.

The results were analysed in terms of the hypotheses outlined earlier, and the stated Probabilities (P) were calculated by The Exact Test of Significance (Fisher, 1936).

### *Brain damage*

The first three brain-damaged subjects to be tested were selected from each of the four groups and compared with a similar selection of subjects who showed no evidence of brain injury. Brain-damaged subjects consisted of seven epileptics, cases of meningitis, encephalitis and cases involving gross motor co-ordination, birth injury, and severe delay in early development. There was found to be no significant difference between the two groups ( $p = 0.207$ ). In fact, the probability represents a tendency for the non-brain damaged to be more distractible than the brain damaged (see later).



Table 1. *Distractibility of Brain-damaged (BD) v. Non-brain damaged (NB) subjects*

	12 BD	v.	12 NB
Not distractible	8		5
Distractible	4		7
	$p = 0.207$		
Details of above Groups:			
Age first institutionalized			
Median	10		5
Inter-quartile range	8-11		1-8.5
Range	0-14		0-17
Length of institutionalization			
Median	7		11
Inter-quartile range	2.5-7		7.5-18.5
Range	1-19		1-25
Age			
Median	17.5		19
Inter-quartile range	15.5-20.5		18-21.5
Range	12-27		16-29
I.Q.			
Median	61.5		67.5
Inter-quartile range	51-72		59.5-77.5
Range	46-79		53-86

*Early environment (see Tables 2a & b)*

The three subjects with earliest institutionalization were selected from each of the four groups and contrasted with 12 subjects who were institutionalized relatively late in life, i.e. the three latest institutionalized subjects from each group. It can be seen that subjects who were institutionalized early in life were also institutionalized for longer periods than the later institutionalized group. Early institutionalization appears to be associated with distractibility ( $p = 0.045$ ), the significance level being enhanced when subjects who had been institutionalized for over 7 years were excluded from the late institutionalized group ( $p = 0.006$ ). When the eight earliest institutionalized were compared with the eight latest the significance level was again increased ( $p = 0.0035$ ), thus it would appear that the earlier the institutionalization the greater the amount of distractibility.

The exclusion of long institutionalized cases from the late institutionalized group, referred to above, enhanced the  $p$ -value, thus suggesting that long institutionalization may have some effect on distractibility. Indeed, subjects hospitalized at an early age tended to be institutionalized for long periods, e.g. the mean period of institutionalization for early and late institutionalized cases was 13.3 years and 4.8 years respectively. However, a comparison of the 12 longest and 12 shortest institutionalized patients\* was insignificant though in the predicted direction ( $p = 0.107$ ). It is of interest that the subjects in these two groups were institutionalized at mean ages of 5.75 and 14.10 years, that is, considerably later than the early institutionalized group mentioned above.

Table 2a. *Distractibility of Early (E) and Late (L) institutionalized subjects*

(Fisher's Exact Significance Test)

	12 E	v.	12 L	12 E	v.	12 L	8 E.	v.	8 L
				(excluding inst. for 7 + years in L group)					
Not distractible	2		7	2		9	0		6
Distractible	10		5	10		3	8		2
	$p = 0.045$			$p = 0.006$			$p = 0.0035$		

Details of institutionalization for the above groups:

Age first institutionalized									
Median	2	17	2	15.5	17	0.5			
Inter-quartile range	0-4.5	14-17.5	0-4.5	12.5-17.5	17-18	0-3			
Range	0-8	10-28	0-8	10-28	14-28	0-5			
Length of institutionalization									
Median	12	3	12	2.5	2	15			
Inter-quartile range	8-20	1-7	8-20	1-5	1-3	10-20			
Range	1-24	0-16	1-24	0-7	0-16	1-23			
Age									
Median	19	20.5	19	18.5	19.5	19			
Inter-quartile range	17-21	18-25	17-21	18-21	18-25	16-21			
Range	13-29	12-33	13-29	12-29	18-33	13-23			
I.Q.									
Median	66	65	66	69	63	60.5			
Inter-quartile range	49.5-77.5	54-77	49.5-77.5	58-77	54-72.5	43.5-73			
Range	32-86	49-87	32-86	52-87	49-87	32-86			



Table 2b. *Distractibility of Long (L) v. Short (S) institutionalized subjects*

(Fisher's Exact Significance Test)

	12 L	v.	12 S
Not distractible	5		9
Distractible	7		3

 $p = 0.107$ 

Details of institutionalization for above group:

Age first institutionalized			
Median	4.5		15.5
Inter-quartile range	0-10		10.5-17.5
Range	0-15		0-28
Length of institutionalization			
Median	18.5		1.5
Inter-quartile range	12.5-21		1-3
Range	11-25		0-7
Age			
Median	22.5		18
Inter-quartile range	20.5-29		16-20.5
Range	18-29		12-29
I.Q.			
Median	58		64
Inter-quartile range	49-71		54-77
Range	41-79		46-87

Analysis of home background and distractibility showed insignificant difference (see Table 3). Detailed histories were not available in every case and these were omitted from the data.

Table 3. *Home conditions and distractibility*

	No home	Very bad home	Less adverse home	Unclassified
Not distractible	0	3	11	8
Distractible	4	5	9	

*Intelligence—Verbal/Performance discrepancies*

It was hypothesized that persons with low Performance relative to Verbal I.Qs might be more distractible than subjects where the reverse occurred. This did not appear to be the case for a comparison of the 12 subjects with the largest  $P > V$  discrepancy with a similar number of subjects with  $V > P$  showed no significant difference in terms of distractibility ( $p = 0.500$ ).

The non-distraction drawings of the 12 most intelligent and 12 least intelligent subjects\* were, however, compared for accuracy and there is clearly a highly significant difference between them ( $p = 0.006$ ). When the same subjects are compared on distraction non-distraction, there is no apparent relationship with I.Q. ( $p = 0.500$ ) (see Table 4). Thus it would appear that distractibility, as measured here, is independent of intelligence.

It is of interest that the two groups also had different median periods of

institutionalization and therefore the results may to some extent contra-indicate any distraction effect being related to length of institutionalization.

### Age

The 12 eldest and 12 youngest subjects\* were compared for distractibility but the results show no significant difference ( $p = 0.500$ ).

Table 4. *Distractibility of Eldest (E) v. Youngest (Y) and Most Intelligent (M) v. Least Intelligent (L) subjects*

	12 E	v.	12 Y	12 M	v.	12 (L)
Not distractible	5		4	5		6
Distractible	7		8	7		6
	$p = 0.500$			$p = 0.500$		
Details of above groups :						
Age first institutionalized						
Median	10		10	9.5		8
Inter-quartile range	6-14.5		5.5-12	5.5-11		1-14
Range	0-28		0-18	0-28		0-17
Length of institutionalization						
Median	13.5		7	8		13
Inter-quartile range	9.5-18		1-8.5	4.5-11.5		2-19.5
Range	1-23		0-12	1-19		1-23
Age						
Median	24.5		16.5	19		19.5
Inter-quartile range	22-29		14.5-17.5	18-21.5		16.5-26
Range	20-33		12-18	12-29		13-29
I.Q.						
Median	60.5		61.5	78		49
Inter-quartile range	48.5-76		53.5-72	70-81.5		46.5-56
Range	41-87		32-84	67-87		32-69

## DISCUSSION

It is of interest that the results of brain-damaged subjects do not confirm those of Strauss & Lehtinen (1960) but even have a non-significant tendency in the opposite direction. Since early institutionalization appears to be a factor in distractibility it may be relevant that the brain-damaged subjects were institutionalized at a later period than the non-brain damaged (see Table 1). Early institutionalization appears to be associated with distractibility, though the rôle of long institutionalization is not clear due to differences in the age of first institutionalization in the group comparisons. However, the results suggest it may be a variable of some importance, possibly as a concomitant of early institutionalization.

Thus, in this research, factors other than brain damage seem important in determining distractibility. This would appear consistent with the criticism of

\* The 12 subjects consisted of the three most extreme members of each of the four groups in terms of the variable being analysed.



Strauss & Lehtinen mentioned earlier in the paper. It seems that although some brain-damaged subjects may appear distractible the brain damage may not be the cause of their distractibility.

That early and possibly protracted institutionalization should be relevant factors in distractibility would appear of interest, particularly in view of recent studies on deprivation. Clarke & Clarke (1958) have discussed various forms of social deprivation, including institutionalization, and have noted their effects on the cognitive performance of subnormals. It would seem possible that fluctuating concentration may be one of the variables involved. In fact, Payne (1960), summarizing research into the cognitive abnormalities of mental patients, suggested that the effect of distraction has been underestimated.

Home conditions were not shown to be associated with distractibility. Two factors may account for this: firstly, it was not possible to analyse all homes in terms of 'very bad' and 'less adverse' environments because of lack of social information, and secondly, although there were a number of homes classed as 'less adverse', many of them were borderline 'very bad' in terms of the criteria. It is thought that the home backgrounds are probably too similar for any differential results to be obtained and before any conclusions can be reached subjects from 'good' home backgrounds should be included.

The above results have been analysed in terms of negative score (distraction) and positive scores (non-distraction). An unexpected result was that not only were certain subjects non-distractible but tended to perform more adequately under visual distraction than with no such stress. These subjects included both brain-damaged and non-brain damaged subjects. They tended to be institutionalized after 7 years of age and have only short periods of institutionalization.

#### SUMMARY

Forty subjects were investigated on a visual distraction task and a number of hypotheses examined.

Brain-damaged subjects were no more distractible than non-brain damaged. Subjects who were institutionalized early in life were significantly more distractible than those first institutionalized at a later date. Relatively late, and perhaps short, institutionalization were associated with improved performance under mild visual distraction or stress.

It was not possible to clearly analyse the relationship between home background and visual distractibility due to lack of sufficient variation between the home environments and inadequate social information, and it is suggested that further investigations are required before any conclusions can be drawn.

Intelligence was not found to be associated with visual distractibility in this experiment though ability to carry out the task under conditions of no distraction was highly correlated with cognitive ability.

Thanks are due to Miss Ethel Duncan, Lecturer in Statistics, University of Bristol, for statistical help and advice; to Dr W. Lumsden Walker, Medical Superintendent, Hortham-Brentry Hospital Group, and Mrs P. M. Brown for their constructive criticisms, and also to Mrs B. Armishaw for secretarial help.

## REFERENCES

- BROWN, R. I. & CLARKE, A. D. B. (1963). The effects of auditory distraction on institutionalised subnormal and severely subnormal persons. *J. ment. Defic. Res.* (In the press).
- CLARKE, A. D. B. & CLARKE, A. M. (1954). Cognitive changes in the feeble-minded. *Brit. J. Psychol.* **45**, 173-179.
- CLARKE, A. M. & CLARKE, A. D. B. (1958). *Mental Deficiency: The Changing Outlook*. London: Methuen.
- FISHER, R. A. (1936). *Statistical Methods of Research Workers*. Oliver & Boyd.
- PAYNE, R. W. (1960). Cognitive abnormalities. In *Handbook of Abnormal Psychology* (ed. Eysenck, H. J.). London: Pitman Medical Publishing Co. Ltd.
- POND, D. A. (1961). Psychiatric aspects of epileptic and brain-damaged children. *Brit. med. J.* **2**, 1377-1382, 1454-1457.
- STRAUSS, A. A. & LEHTINEN, L. E. (1960). *Psychopathology and Education of the Brain Injured Child*. New York: Grune & Stratton.
- WOODWARD, Mary (1960). Early experiences and later social responses of severely subnormal children. *Brit. J. med. Psychol.* **33**, 123-132.

*Manuscript received 20 December 1962*



## Measures of Distance and Motility in Psychotic Children and Severely Subnormal Controls

BY N. O'CONNOR AND BEATE HERMELIN

*Medical Research Council, Social Psychiatry Research Unit,  
Institute of Psychiatry, London, S.E.5*

In a group of severely subnormal children diagnosed as psychotic, and controls, the approach response and the amount of movement were measured in the presence of various stimuli. An attempt was made to isolate relevant aspects of 'a person' in preparing the stimuli.

The distance measure did not differentiate the groups in those conditions which resembled most closely a social stimulus situation. They did differ in their approach score to objects, the psychotics approaching less. Psychotic children also moved more than controls, and the amount of movement did not vary over different stimulus situations.

### INTRODUCTION

The literature on psychotic or autistic children, as the second descriptive adjective implies, frequently reports their lack of interest in human or social contacts. Kanner (1948) and Creak *et al.* (1961) are two cases in point. However, a previous experiment by Hermelin & O'Connor (1963) showed that a social response measure did not differentiate psychotic and control imbeciles except in so far as speech was concerned. The present experiment was designed to examine some aspects of social stimulation which might determine the degree of responsiveness.

Harlow (1962) has demonstrated that it is possible to isolate the relevant aspects of a stimulus which evoke clinging responses in young monkeys, normally directed towards the mother. In our experiment we attempted to determine whether psychotic children would show differential approach and withdrawal responses to isolated aspects of 'a person', and whether they would differ in this from severely subnormal controls. The selected stimulus aspects were softness, rocking movement, visual likeness and voice. Two control stimuli were also included. These were a neutral object and one of the experimenters (B.H.).

### METHOD

#### *Subjects*

Twelve institutional psychotic imbecile children were selected by a psychiatrist on the basis of the presence of manneristic or autistic behaviour. Controls were matched for age, sex and I.Q. Each subject was then checked against Creak *et al.*'s (1961) list of symptoms. No child was accepted for the psychotic group unless he had at least three of the symptoms, and the average number of symptoms present in the psychotics was four. No controls showed any symptoms. The mean age of the control group was 14 years 2 months, range 11 to 16 years, and of the psychotic group 13 years 6 months, range 9 to 16 years. The mean I.Q.s (Wechsler Adult Scale) were 40 points, range 18 to 58, and 36.5 points, range 25 to 50 points respectively. It should be added that I.Q. estimation with psychotics is highly unreliable. Clinical diagnoses for all subjects were obtained from the case records. Table 1 summarizes these. The table also gives comparative data for a group of patients in another hospital. These patients took part in a similar experiment and the two sets of data can therefore be compared (Hermelin & O'Connor, 1963).

Table 1. *Incidence of relevant clinical signs*

Groups	Hospitals	Row	Complications of pregnancy and delivery toxaemia etc.	Encephalitis and meningitis	Epilepsy	Strabismus	Overt neurological disease
Psychotics	{ 1 2	1	4	3	3	3	1
		2	7	0	1	4	3
			11	3	4	7	4
Controls	{ 1 2	3	7	0	0	4	2
		4	0	0	3	0	1
			7	0	3	4	3



Complications of pregnancy include toxæmia, prematurity and asphyxia neonatorum, and overt neurological disease includes one case of gargoylism and is otherwise made up of cerebral palsy and hemiplegia. The table summarizes data of considerable relevance to disturbed behaviour but of very varying reliability. A comparison of hospitals 1 and 2 within each group shows considerable variation especially in the control group. This means that no undue weight can be placed on the results. However, by comparing groups within the first hospital a total of 14 and 13 relevant signs are recorded (Rows 1 and 3). These figures suggest that psychotic children suffer as many clinical anomalies as the control imbeciles. A comparison within the second hospital (Rows 2 and 4) shows row totals of 15 and 4 respectively. In spite of the unreliability of the Row 4 total these results do not contradict the conclusion from the first figures that at least as many psychotic children as controls show indications of gross cerebral lesions. If total figures are compared, three items are implicated in any possible group difference, complications of pregnancy, encephalitis-meningitis and strabismus. Attention should perhaps be paid to the associations between 'psychosis' in children and these signs.

### Procedure

Each subject was tested in six conditions, each lasting for 5 minutes. Order of condition between subjects was determined by a 6 x 6 Latin Square repeated twice for each group. One condition (A) was the presence of a large square cardboard box (control stimulus); another (B) was the presence in the same place of a blanket hung on a wire between the walls (softness); a third (C) was a small rocking platform placed on the floor in the same place (rocking movement); the fourth (D) was a realistic life-sized model of a woman on a chair (visual likeness), and the fifth (E) a loudspeaker placed at the same end of the room from which the experimenter's voice could be transmitted (voice). A sixth condition (F) was the presence in the room of one of the experimenters sitting on a chair. All stimuli were placed at the end of a room of one of the experimenters sitting on a chair. The window was completely blacked out, and the room 12 ft x 8 ft room, opposite the door. The window was completely blacked out, and the room was illuminated by an overhead 40 watt bulb. On one wall was a large one-way screen. The subject was brought into the room and asked to remain there for 5 minutes. For each subject in three 5-minute sessions took place in the morning and three approximately 1½ hours later in the afternoon. Otherwise there was a time interval of ½ hour between sessions.

Data were recorded in the following manner: at increasing distances from the display three lights shining on three photo-electric cells let into the skirting board activated three pens on a recorder in an adjoining room. The distance between cells was 3 ft. Whenever a subject crossed between a light and a photo cell on the opposite wall, the light beam was interrupted and this interruption was marked on the recorder. Thus the subject's position in any one of four parts of the room was noted continuously. Distance was scored as the time spent by each subject in each part of the room. In the section nearest the display this time score was weighted by a factor of 4, in the next removed by 3, in the next by 2 and in the last by 1. These weighted scores were summed to give the time-distance score for any one session or condition. In addition a motility score was calculated. This was a simple addition of the number of times a light was crossed in any session as shown by the marker pens.

## RESULTS

Mean weighted distance and motility scores for each group in each of the six conditions are given in Table 2. In the distance data high scores reflect a close approach to the stimulus.

Table 2. Mean weighted distance and movement scores

Groups	Conditions					
	A	B	C	D	E	F
Distance scores						
Psychotics	613 ± 200	569 ± 261	568 ± 203	616 ± 273	572 ± 222	842 ± 259
Controls	758 ± 274	718 ± 274	727 ± 273	600 ± 270	604 ± 222	922 ± 220
Movement scores						
Psychotics	28 ± 34	26 ± 40	44 ± 48	33 ± 47	21 ± 22	42 ± 59
Controls	13 ± 10	15 ± 16	15 ± 18	7 ± 5	12 ± 13	24 ± 24

Separate analyses of variance for distance and motility scores were carried out. Dealing first with the distance data, as can be seen from Table 3 there was a significant difference between groups. Subsequent *t*-tests showed that it was conditions A, B and C and not D, E and F which contributed significantly to this difference. Psychotics remained more distant than controls in the first three conditions and did not differ significantly from them in the latter.

Table 3. *Analysis of variance table—Distance data*

Source	ds.f.	$v^2$	F	P
Between groups	1	303,417.29	7.04	0.01
Psychotic group				
Between subjects	11	226,328.59		
Between orders	5	88,139.79	2.04	N.S.
Control group				
Between subjects	11	196,243.06		
Between orders	5	162,648.73	3.77	0.01
Between conditions	5	272,090.37	6.30	0.001
Groups $\times$ conditions	5	31,095.37		N.S.
Residual	100	43,113.60		

There was also a significant difference between conditions. This was due to the fact that all subjects approached significantly more in condition F, the presence of one of the experimenters, than in any other condition. There was no significant interaction.

Table 4. *Analysis of variance table—Motility data*

Source	ds.f.	$v^2$	F	P
Between groups	1	11,610.06	20.52	0.001
Psychotic group				
Between subjects	11			
Between orders	5	400.9	0.71	N.S.
Control group				
Between subjects	11			
Between orders	5	118.89	0.21	N.S.
Between conditions	5	937.59	1.66	N.S.
Groups by conditions	5	391.2	0.69	N.S.
Residual	100	564.67		

Concerning Motility scores (Table 4), there was no difference between conditions. However, there was a clear difference between groups. Psychotics moved about more than controls. But though psychotics have higher motility scores than controls, there is also a relationship between motility and I.Q. As subjects in the two groups were individually matched for I.Q., the I.Q. range within the groups made it possible to correlate I.Q. with the two main scores. These correlations were negative, that with the distance data being  $-0.12$ ,  $p = \text{N.S.}$ , and that with the motility data  $-0.57$ ,  $p = 0.01$ . Amount of motility thus correlated significantly with low I.Q.

#### DISCUSSION

The aim of this experiment was to isolate those aspects of social contact to which children described as psychotic or autistic would respond. This aim has not been



realized in this experiment. A plausible interpretation of the results would be that the children reacted to the blanket and the rocking platform in a manner similar to that in which they reacted to the neutral object, i.e. the box. Indeed, the children could probably be described as having been more reasonable than the experimenters, who somewhat naïvely hoped that these former two conditions might represent personal attributes of softness and movement to the children.

On the other hand, the female model and the voice may have been more readily recognized as human attributes. However, even these stimuli did not lead to responses in the children such as Harlow's cloth mother evoked in the monkeys. This is emphasized by the significant increase in approach scores for all children, when the display consisted of one of the experimenters.

This striking response in all cases confirms the finding of a previous experiment (Hermelin & O'Connor, 1963) that autistic children of this intelligence level respond positively to an adult. These findings must be held to qualify clinical observations that psychotic children are socially withdrawn. An occasionally stated obverse of the supposed lack of interest in adults is the preoccupation of autistic children with objects. Neither the previous experiment in which such children played less with toys, nor this in which they approached relatively inanimate objects (A, B and C) less than controls would appear to support this view.

If we finally consider the amount of general motility shown by the two groups, the experiment clearly indicates that psychotics show more general motility, just as in the previous experiment they showed more arm and hand movements. The fact that this greater motility was not detected in the first study must be accounted for in terms of the improved continuous recording technique used on this occasion. As motility remained constant over all conditions it must be regarded as independent of external stimuli. It may not, however, be independent of I.Q. An association exists between low I.Q. and high motility. As progressively more cortical damage is found with decreasing I.Q.s, it may be the case that in the present sample, low I.Q., increased motility and autism, may have a common foundation in more marked and extensive central nervous system anomalies.

#### Acknowledgements

We wish to acknowledge the aid of Dr B. W. Richards and Dr P. E. Sylvester who helped with diagnoses, and the facilities provided by Dr J. Gibson at St Lawrence's Hospital, Caterham. We would also like to thank Dr P. H. Venables for statistical advice.

#### REFERENCES

- CREAK, MILDRED, *et al.* (1961). Schizophrenic syndrome in children. *Lancet*, **ii**, 818.  
HARLOW, H. F. (1962). The development of affectional patterns in infant monkeys. In *Determinants of Infant Behaviour* (ed. B. M. Foss). London: Methuen.  
HERMELIN, BEATE & O'CONNOR, N. (1963). The response and self-generated behaviour of severely disturbed children and severely subnormal controls. *Brit. J. soc. clin. Psychol.* **2**, 37-43.  
KANNER, C. (1948). *Child Psychiatry*, (2nd edn.). Oxford: Blackwell.

## Speed and Accuracy on Mazes in Relation to Diagnosis and Personality

By G. A. FOULDS AND ANNA OWEN

*Runwell Hospital, Wickford, Essex*

### INTRODUCTION

The *Wrong Directions (WD) measure* on the Porteus Mazes has long been regarded as a measure of intelligence, involving especially foresight (Porteus, 1942). In a comparison between schizophrenics and neurotics, matched for Vocabulary level, one might expect that schizophrenics would show evidence of some impairment. Since no difference would be expected in ability to thread mazes as between Hysteroid and Obsessoid personalities, schizophrenics should have a higher WD score, regardless of personality, than should neurotics. There should, therefore, be a positive correlation between WD and a symptom measure which differentiates between schizophrenics and neurotics.

Since the *Starting Time measure (ST)*—the time elapsing between *S* being instructed to start and his actually starting to trace (Foulds, 1951)—relates, not to the ability to do the test, but rather to the manner of doing it, this might be expected to relate to personality characteristics more than to diagnosis. Obsessoid personalities should, therefore, have a higher ST score than Hysteroid personalities, regardless of their diagnosis. The Hysteroid score (on the HOQ) should, therefore, correlate negatively with the ST score.

### PROCEDURE

#### *Subjects*

The intention was to include 25 successive female admissions diagnosed by psychiatrists as schizophrenic and 25 as neurotic. One subject, originally diagnosed as neurotic, was re-diagnosed as schizophrenic, leaving 24 neurotics. Of the 26 schizophrenics, one failed to reach the criterion level of 33 on the Mill Hill Vocabulary. Five schizophrenics failed to meet the criterion set for valid performance on the Mazes, leaving 20 schizophrenics. All were acute cases in the sense of having spent less than 2 years in hospital altogether. All were tested within 1 week of admission; but this was not always the first admission.

#### *Tests*

1. Mill Hill Vocabulary Scale; 2. The Non-integrated Psychotic Scale from the Runwell Symptom-Sign Inventory; 3. The Hysteroid-Obsessoid Questionnaire—HOQ (Caine & Hawkins, 1963); 4. The Porteus Mazes.

In view of the known difficulty experienced by schizophrenics in comprehending the Mazes, the administration was modified. The same order was kept; but Year 5 was demonstrated by *E* and Years 6, 8 and 11 were trial runs for *S*, in which errors were pointed out. Only Years 7, 9, 10, 12 and 14 were scored. The ST measure was the square root of the total for the 5 mazes of seconds elapsing between *E*'s saying 'go' and the beginning of actual tracing by *S*.

The Non-integrated Psychotic Scale (NIP) consists of those items from the Runwell Symptom-Sign Inventory which had been found to differentiate schizophrenics from other psychotic and from all neurotic subgroups (Foulds & Owen, 1963).

### RESULTS

Eighteen of the 20 psychiatrically diagnosed schizophrenics fell above the cutting score previously established for the NIP Scale against 4 of the 24 neurotics. This scale was therefore used as the diagnostic measure (cube root transformation).

Table 1 shows the product-moment correlations between the six measures.



Table 1. *Product-moment correlations between the six measures*

HOQ	ST	WD	AGE	MHV	
0.212	-0.170	0.307*	-0.174	-0.089	NIP
	-0.412†	0.115	-0.136	-0.303*	HOQ
		-0.049	0.314*	0.064	ST
			0.123	-0.160	WD
				0.177	AGE

\*  $p < 0.05$ . †  $p < 0.01$ .

There is a significant tendency for those who score towards the schizophrenia end of the NIP Scale to have more Wrong Directions on the Mazes, regardless of personality type.

There is a significant tendency for those who score towards the Hysteroid end of the HOQ to start tracing the Mazes more quickly, regardless of diagnosis. This relationship is little affected when age is held constant ( $-0.393$ ). It would appear, therefore, that the HOQ measures something meaningful in acute schizophrenics, though this might not hold true of chronic schizophrenics.

Older Ss start tracing more slowly and Obsessoid personalities tend to have somewhat better vocabularies.

This study forms part of projects supported by a grant from the Medical Research Council.

## REFERENCES

- CAINE, T. M. & HAWKINS, L. G. (1963). Questionnaire measure of the hysteroid/obsessoid component of personality. *J. cons. Psychol.* **27**, 206-209.
- FOULDS, G. A. & OWEN, ANNA (1963). Are paranoids schizophrenic? *Brit. J. Psychiat.* **109**, 674-679.
- FOULDS, G. A. (1951). Temperamental differences in Maze performance. *Brit. J. Psychol.* **42**, 209-217.
- PORTEUS, S. D. (1942). *Qualitative Performance on the Maze Test*. Vineland, N. J., Smith.

*Manuscript received 30 May 1963*

## A Comparison of Four Methods of Scoring an Attitude Scale in Relation to its Reliability and Validity

BY PAMELA K. POPPLETON AND G. W. PILKINGTON

*Department of Education, University of Sheffield, and Department of Psychology, University of Sheffield*

The responses of 120 subjects to two parallel versions of a religious attitude scale were scored by four different methods: Thurstone scoring; Likert scoring; scale-product scoring; and scoring by weighted proportions. The effect of these methods on the reliability and validity of the scale was then compared.

Likert scoring and scoring by weighted proportions produced the highest reliability coefficients (0.95), and Thurstone scoring the lowest (0.85). Scale-Product scoring resulted in a reliability coefficient of 0.88. Reasons are given for considering this method of scoring to be unsatisfactory.

The validity coefficients given by each of five criteria over the four different methods of scoring were found to be quite closely comparable, but generally scoring by weighted proportions yielded the highest validity coefficients. But the simpler Likert system did almost as well.

### A. INTRODUCTION

In the course of a study previously reported (Poppleton & Pilkington, 1963), two parallel forms of a scale for the measurement of religious attitudes were constructed using the Thurstone method for the compilation of statements. They were used in a preliminary survey amongst a group of 120 subjects before conducting an item analysis and drawing up the final version of the scale. Each form consisted of 22 items, subjects being asked to respond to each item by endorsing one of five categories: 'Strongly Agree', 'Agree', 'Uncertain', 'Disagree' and 'Strongly Disagree'. Half the group had Form A first, and half Form B, the other form being given three weeks later. They were scored by four different methods:

1. Ordinary Thurstone scoring on items which were endorsed.
2. Likert scoring. Response categories were weighted 5-4-3-2-1-, and the weights reversed at the mid-point of the scale.
3. The Scale-Product method. A modified version used by Castle (1953) was adopted, in which the Likert weight is multiplied by the deviation of the Thurstone scale value of the item from the mid-point of the scale (5.5).
4. A method suggested by Guilford (1954) which uses empirically derived weights for the response categories, and which will be referred to here as 'the method of weighted proportions'.

### B. RESULTS AND DISCUSSION OF RESULTS

Data concerning reliability and validity are shown in Tables 1 and 2.

#### *Reliability*

A Thurstone score based on the median position of each respondent is essentially a limen score indicating the subject's central response tendency towards the attitude.



Table 1. *Statistics on four methods of scoring parallel forms*

	Form A	Form B	Product moment correlation $N = 120$
Thurstone Scoring			
Mean	4.9	4.6	} 0.85
S.D.	1.7	2.0	
Range	7.24	7.33	
Likert Scoring			
Mean	63.0	66.4	} 0.95
S.D.	14.3	13.4	
Range	60	59	
Scale Product Method			
Mean	14.1	23.4	} 0.88
S.D.	47.8	45.4	
Range	206	199	
Method of Weighted Proportions			
Mean	87.1	90	} 0.95
S.D.	20.5	19.6	
Range	74	74	

But variability is introduced because respondents may endorse as few as three or as many as 14 items, so that the size of the correlation between limen scores tends to be reduced. The Likert method of scoring by summing the responses in each category has commonly been found to yield higher reliability coefficients. For example, Mosier (1941) investigated the relationship between the limen score and the summation score of a mental test, and found the reliability of the first to be 0.88 compared with 0.94 for the second. The coefficients yielded here by our attitude scale are closely comparable.

Table 2. *Validity.  $r_{bis}$  between scale scores and reported religious activities (Form B)*

Standard error $r_{bis}$ shown in brackets					
Method of scoring (N = 120)					
Criterion	Thurstone	Likert	Scale-Product	Weighted proportions	
(a) Active church membership	0.84 (0.05)	0.86 (0.05)	0.84 (0.05)	0.88 (0.04)	49
(b) Church attendance (twice or more) in a month	0.74 (0.06)	0.79 (0.06)	0.81 (0.06)	0.86 (0.05)	44
(c) Daily private prayers	0.72 (0.07)	0.73 (0.07)	0.73 (0.07)	0.72 (0.07)	28
(d) Membership of a student religious group	0.60 (0.09)	0.55 (0.09)	0.57 (0.09)	0.60 (0.09)	32
(e) Expressing some form of religious belief	0.93 (0.03)	0.96 (0.02)	0.94 (0.03)	0.97 (0.02)	63

The Scale-Product method represents the major attempt to combine the Thurstone and Likert methods. The results reported, however, do not support Butcher's (1956) hypothesis that this method gives a higher reliability coefficient

than Likert scoring before item analysis has been applied. Our reliability coefficient of 0.88 may be compared with Butcher's (0.89) and Castles' (0.81), although Eysenck and Crown (1949) quote one of 0.94 for their Anti-Semitism scale. But Butcher presents a detailed criticism of their use of the Scale-Product method, suggesting that the weights actually used are difficult to reconcile with their description of the method. It appears that their weights approximate closely to those achieved by the method of weighted proportions, so that the resulting scores cannot really be called 'scale products'. The lower reliability of true Scale-Product scoring can probably be explained by the fact that the Thurstone scale values of items in parallel forms are close but not identical, and multiplying them by response weights exaggerates the differences. Table 3 shows four pairs of items with approximately the same Thurstone scale values. The deviation of each of these from the mid-point of the scale is multiplied by five to arrive at the weight for the 'Strongly Agree' category.

Table 3. *Thurstone scale values of parallel items and their derived weights*

Form A			Form B		
Thurstone Scale value of item	Deviation from mid-point of scale (5.5)	Weight for 'Strongly Agree' category	Thurstone Scale value of item	Deviation from mid-point of scale (5.5)	Weight for 'Strongly Agree' category
3.15	2.35	12	2.90	2.6	13
3.64	1.86	9	3.32	2.18	11
3.97	1.53	8	3.87	1.63	8
4.41	1.09	5	4.22	1.28	6

The two sets of items may be regarded as parallel to all intents and purposes, but they nevertheless receive different weights on the two forms. Adding these differences means that subjects who obtain identical scores on the Likert system can be separated by as much as ten or eleven points using scale-product scores. It would seem, therefore, that this method can never give a high coefficient of equivalence because of its tendency to exaggerate small differences in the original scale values. It is as if one were taking measurements with a piece of elastic fastened at the middle.

An alternative method of retaining information about extremeness and intensity is to use weights derived from the proportion of respondents who endorse an item in each of the five categories. Guilford (1954) gives an account of such a method. The summation scores are divided at the mean, and the proportion of high and low scores for each item in each category is calculated. Weights are then assigned by entering the proportions in an abac given by Guilford (p. 446). This was the procedure we followed. Eysenck (1954) suggests that this method may prove more reliable than the Scale-Product method, but this has not until now been tested. Our results show a very satisfactory reliability coefficient, although it is still no higher than that yielded by Likert scoring.

#### *Validity*

Table 2 shows that the correlation coefficients given by each criterion for the four



different methods of scoring are quite closely comparable. There is no one 'best order', but on practically all indices the method of weighted proportions yields the optimum weights for high validity. Weighting has been shown to be important in relation to validity when the number of items is not large (Guilford, 1954; Gulliksen, 1958). But when, as with our scale, there are over 20 items, Guilford suggests that weights can depart from their optimal values without seriously affecting validity. Moreover, the different weighting systems could not, in principle, make much difference, because all used positive values in the same directions, and so might be expected to yield comparable results.

### C. CONCLUSIONS

This study upholds the general finding that summation scores give higher reliability coefficients than limen scores. The Likert method is also found to provide as high a reliability as the method of weighted proportions. The Scale-Product method, however, when used as intended, does not give as high a reliability as the latter two methods. It is argued that this type of scoring is unsatisfactory. The method of weighted proportions also provides a good indication of validity. But when the number of items is more than 20 the system of weights used does not affect validity seriously. In general, the method of weighted proportions has much to recommend it, but the simpler Likert system does nearly as well.

### REFERENCES

- BUTCHER, H. J. (1956). A note on the scale product and related methods of scoring attitude scales. *Brit. J. Psychol.* **47**, 133-139.
- CASTLE, P. F. C. (1953). A note on the scale-product technique of attitude scale construction. *Occup. Psychol.* **27**, 104-109.
- EYSENCK, H. J. (1954). *The Psychology of Politics*. London: Routledge & Kegan Paul.
- EYSENCK, H. J. & CROWN, S. (1949). Primary social attitudes: II. An experimental study in opinion-attitude methodology. *Internat. J. Opin. Attit. Res.* **3**, 47-86.
- GUILFORD, J. P. (1954). *Psychometric Methods*, 2nd edn. London: McGraw-Hill.
- GULLIKSEN, H. (1958). *Theory of Mental Tests*. New York: Wiley.
- MOSIER, C. (1941). Psychophysics and mental test theory. *Psychol. Rev.* **48**, 235-249.
- POPPLTON, P. K. & PILKINGTON, G. W. (1963). The measurement of religious attitudes in a university population. *Brit. J. soc. clin. Psychol.* **2**, 20-36.

*Manuscript received 4 November 1962*

## **Salience of Attributes and Commitment to Extreme Judgments in the Perception of People**

BY HENRI TAJFEL AND A. L. WILKES\*

*Department of Social and Administrative Studies, Oxford University*

The investigation is concerned with the relation between the subjective importance of a personal attribute and the manner of judging other people in terms of that attribute. A prediction was made that in a task involving the rating of others in terms of a number of attributes, there will be a tendency to use relatively extreme judgments for attributes which are important to the person making the judgments. This hypothesis, derived from a background of general findings about quantitative judgment, was confirmed. It was tested in a situation in which the subjects were able to give fairly free descriptions of others. Some time after these descriptions were made, each subject rated photographs on rating scales based on his own use of various attributes. The salience of attributes (i.e. their relative priority and frequency of occurrence in free descriptions of others) was assumed to be an index of their importance to the subjects. This assumption was tested and validated in a separate experiment.

### 1. INTRODUCTION

Experimental work on perception of people is beset with difficulties inherent in the nature of the problems studied. These investigations attempt to extend the scope of generalizations that can be made about the ways in which people perceive, assess or judge other people. Reliable generalizations are, of course, most easily established in properly controlled experiments; but adequate controls imposed on the fleeting, variable and rich contexts of impressions about people have a way of yielding predictions which seem to apply above all to other equally well-controlled situations. The main reason for the existence of this vicious circle of prediction from experiment to experiment instead of prediction from experiment to natural phenomenon is probably in the difficulty (or sometimes impossibility) of identifying the relevant determining variables in 'real' situations.

The size of this problem is such that no single solution to it seems possible. The most one can hope for is an increase in the number of experiments which have a dual line of parentage: one relating them to a background of general psychological theory; and another linked to some fairly direct implications reaching outside the laboratory.

The present experiments were designed as one attempt in this direction. The hypotheses were based on a fairly well documented set of findings in the field of quantitative judgment; the results seem applicable without undue qualifications to a variety of settings in which people assess other people.

The investigation of shifts and biases in the assessments of people has been one of the continuing trends of research in this field. This led to the use of rating scales

\* Now in the Department of Psychology, University of St. Andrews, Queen's College, Dundee.



of various sorts for eliciting biases which occur in the assessments of one individual by another, or of one specified group of people by other specified groups. The use of rating scales implies that problems can be conceptualized in terms of laws of quantitative judgment, and therefore a possibility of searching for common principles which may underlie shifts and biases of judgment both in the 'perception' of people and in the more conventional psycho-physical situations. The common characteristics of the two types of judgment become perhaps most suggestive when one considers the intermediate case of assessment of physical attributes of other people: '... an attribute assigned to an individual ... does not float unanchored in a universe of absolute assertions. A man is not 'tall' in an absolute sense; he is *taller* than other people. ... Attributes assigned to people are arranged on a continuum from more to less, and have very little meaning outside their comparative framework' (Tajfel, 1962, p. 35).

The aim of the investigations reported here was to explore the possibility of extending to judgment of human attributes a group of converging results reported in other areas of quantitative judgment. The common denominator of these findings is that a 'polarization' of judgments occurs when the dimensions along which these judgments are made are of high relevance to the subject. Another way to describe the phenomenon is to state that when the differences between the stimuli judged by the subject have in the past acquired some form of emotional or 'value' significance for him, he will tend to use in his estimates of magnitude a relatively high proportion of extreme judgments, placed at one or at both ends of his response scale.

The theoretical aspects of this polarization of judgments were discussed at length in previous publications (Sherif & Hovland, 1961; Tajfel, 1959), and their explanation sought in the acquisition of anchoring properties by stimuli which, for one reason or another, become salient to the subject. The tendency towards polarization was shown to exist in a variety of situations, ranging from absolute judgments of physical magnitudes to placement of items on attitude scales. In one of these experiments (Tajfel, 1959) the subjects judged a series of weights: in two of the experimental situations judgments were preceded by several presentations of the series during which each appearance of the heaviest or of the lightest stimuli of the series was associated with the presentation of a small reward. A comparison of judgments subsequent to these arrangements with judgments in several control situations showed that in the 'value' situations the differences between the estimates of weight of the extremest stimuli of the series were greater. A parallel finding emerged from an experiment (Tajfel & Cawasjee, 1959) in which differences between estimates of size of two coins (2s. and 2s. 6d.) were compared with the corresponding differences between control coins, highly similar to the first pair but presenting no differences in value. The differences between the size estimates of the current coins were considerably and significantly greater than for the control stimuli. The same two pairs of stimuli were presented subsequently to subjects in the United States, unfamiliar with British coins; there was no trace of an accentuation of differences for the first pair (unpublished data).

Pettigrew, Allport & Barnett (1958) reported equivalent findings in their South African study of identifications of race in a perceptual task involving a resolution



of binocular conflict. They referred to the phenomenon as 'a bifurcating tendency in the judgments of highly involved subjects' (p. 276). The task was performed by subjects from the various ethnic and racial groups in South Africa; it consisted of identifying race in pairs of photographs presented stereoscopically. The pairs included all the possible combinations of South African racial and ethnic groups. The Afrikaner subjects had 'a larger percentage of *African* and a smaller percentage of *Coloured* judgments than any other group' (p. 276). Pettigrew *et al.* assume that 'like Hovland & Sherif's subjects, (the Afrikaners) may be yielding black-white, either-or judgments more than other groups because of their greater involvement' (p. 276).

The common feature of all these findings is the polarization in the judgments of *physical* magnitudes under condition of some degree of involvement with the 'series', whether of weights, of coins, or of faces of differing shades. Hovland & Sherif (1952) reported similar results about the placement of series of statements concerning a controversial social issue. They found that groups of subjects who were highly involved in the desegregation issue in the United States tended to place more of the statements at the extremes of a prescribed category scale than did less involved subjects. The phenomenon was particularly marked for statements which expressed opinions strongly at variance with those held by the subject. Manis (1960) described the same phenomenon for groups of students in relation to the issue of fraternities in American colleges. The mean standard deviations of judgments made by students who held strong views for or against fraternities were greater than of judgments made by students who did not feel strongly about the issue.

The determining condition of the findings summarized so far seems to be the subject's involvement with his judgments on a particular dimension. Different degrees of such involvement have an obvious parallel in the perception of people. We assess other people in terms of a very large number of attributes; each of these attributes can be conceived as a 'dimension' varying from 'more' to 'less', e.g. from 'intelligent' or 'kind' or 'honest' to their opposites. The relative importance of the various attributes is bound to vary from individual to individual; e.g., to one person intelligence may be more important than kindness, to another honesty more important than 'strength of character'. A simple inference from the variety of findings discussed previously would be that judgments of subjectively important attributes should tend to cluster more in the regions of extreme responses than judgments of attributes which are less important to the individual.

However, this inference loses much of its simplicity when one considers two stumbling blocks in the way of subjecting it to some sort of valid empirical test. In the first place, it is by no means easy to ascertain the relative importance of various attributes to an individual. Secondly, terms such as 'intelligent' or 'kind' or 'honest' used by one person are not necessarily equivalent in their connotations to the same terms used by another. 'Intelligence' or other such attributes cannot *by definition* be specified in *a priori* objective terms when one is concerned with the subjective aspects of their use, with the private connotations which determine the extent of their salience for an individual.



The investigation described here was an attempt to circumvent these difficulties in the testing of the prediction just stated. The main preoccupation in the designing and conducting of the experiments was to isolate the relevant variables without at the same time distorting them beyond recognition through rigidity of the experimental controls. This was attempted in the following successive stages:

- (i) Identifying and isolating for each subject an equal number of salient and non-salient attributes *defined in his own terms*.
- (ii) Designing a rating situation in which each subject was able to use his own previously identified salient and non-salient attributes.
- (iii) Adopting an assumption concerning certain differences in the free use of attributes which differ in the degree of their importance to the subjects.
- (iv) Validation of the assumption just referred to.

The procedures will be described in detail in the next section. The assumption underlying them is closely related to a recent discussion by Hastorf, Richardson & Dornbusch (1958). One of the main points put forward by Hastorf *et al.* is that it is crucially important 'to study the qualities of a person's experience of others in terms of the verbal categories he uses in reporting that experience' (p. 56). They assume that in such a study the frequency and the sequence of the categories used by a subject may become fairly important as indices of his manner of perceiving others. A more specific assumption related to the problem at hand was made for the purposes of the present study: that when categories of description of others are used freely by the subjects, those that are related to attributes which are subjectively important will tend to be used earlier in a sequence and more frequently than categories related to the less important attributes.

The hypothesis of the study can therefore be rephrased in the following terms: the most frequent and earliest categories used in free descriptions of others will tend to be related in a subsequent rating task to more extreme judgments than the less frequent and later appearing categories of description. The dual task of the experiments was to test this hypothesis and the validity of the assumption linking it to the problem of subjective importance.

## 2. METHOD

### (i) *Free use of categories*

The aim of the first phase of the experiments was to elicit fairly free descriptions of other people. Ten photographs (heads only) of youngish men (approximately between the ages of 20 and 30) were selected from various sources and reproduced on slides. The group of subjects consisted of 17 young men (students at Oxford University and at Ruskin College). Each subject was provided with ten sheets of paper. The sheets were numbered from 1 to 10, one for each photograph, and each sheet was divided horizontally into ten spaces, also numbered from 1 to 10. The subjects were told that their task would be to describe the personal characteristics of each of the men whose photographs would appear on the screen; they should try to find, as far as they could, ten items of description for each photograph, to be written on the sheet in the order of their occurrence; they should describe each characteristic in their own words, using adjectives, short sentences or any other form of short description.

Following these instructions, the photographs were presented on the screen one after another. Each photograph remained on the screen until all the subjects completed the work on the sheet corresponding to it.



*(ii) Determination of salience of attributes*

Not all the subjects were able to specify as many as ten characteristics for each of the photographs. However, the work of each subject yielded a total of between 50 and 100 separate items of description. These items ranged from single adjectives or nouns to short sentences.

The descriptions provided by each subject were then coded by two judges working together. The aim of this coding was to determine for each subject the clusters of expressions which (a) were identical, synonymous or closely related; (b) overlapped as little as possible with other clusters of expressions used by that subject. In this way, the original number of items supplied by each subject was reduced to about ten to fifteen clusters.

Each of the clusters was then evaluated from the original response sheets in terms of its priority for each of the photographs and of its overall frequency of occurrence. An item appearing first in a description of a photograph was assigned a score of 10; the second item a score of 9, and so on, until the last item on the sheet was reached. Therefore, the final score for each cluster consisted of the sum of the ranks that the various items belonging to it were given each time one of them appeared on the sheets. The scores were thus determined by a combination of priority and frequency of occurrence of the items belonging to each cluster. A high score represented a salient attribute; a low score a non-salient one. The attributes used by each subject were ranked for salience from the highest to the lowest.

*(iii) Preparation of individual rating scales for the subjects*

For each subject, the top four and the bottom four attributes were selected from his ranking order. Each of these attributes was then arranged in a bipolar way (e.g. intelligent—unintelligent). In this, the subject's own manner of describing each attribute was used (e.g., one subject may have used descriptions such as strong and weak, another forceful personality and weak personality).

Ten small booklets were then prepared for each subject. Each booklet consisted of eight pages. On each page, one bipolar attribute was presented in the form of a non-verbal seven-point scale, consisting of the two opposites separated by seven clearly spaced positions. In this way, each subject was provided with ten booklets each consisting of eight rating scales, four for his salient attributes and four for the non-salient ones. The sequence of the attributes and the position of the positive and negative opposites were randomized for each booklet.

*(iv) Rating of the attributes*

The same subjects participated in a second experimental session which took place about 4 weeks after the first. Sixteen out of the original 17 subjects were still available. The stimulus material consisted of ten new photographs of young men obtained in the same way as the photographs for the first session. The subjects were requested to rate each photograph on the eight rating scales, using one booklet per photograph. The rating was done by placing a tick above the appropriate position. The ten photographs were presented one by one on a screen, each remaining on the screen until all the subjects completed all the ratings in the appropriate booklet.

To summarize the procedure: the first experimental session provided descriptions of personality characteristics from photographs. The salience for each subject of the attributes he had used was determined from these descriptions. Rating scales for the four most salient and the four least salient attributes were prepared separately for each subject. In the second experimental session, the subjects rated ten new photographs, each on eight rating scales representing each subject's own salient and non-salient attributes.

### 3. RESULTS

The seven points on the rating scale were assigned numbers from  $-3$  for the most unfavourable rating through 0 for the 'neutral' one to  $+3$  for the most favourable one. The prediction that the subjects would tend to use more extreme ratings for the salient attributes than for the non-salient ones could thus be tested by comparing



the frequency of ratings assigned by each subject to the categories  $-3$ ,  $-2$ ,  $+2$  and  $+3$  for each of the two groups of four attributes.

The general pattern of results is presented in Table 1. An inspection of this table will reveal two trends in the data:

(i) The frequency of the more extreme ratings ( $+3$ ,  $+2$ ,  $-2$ ,  $-3$ ) is consistently higher (with one small exception at  $+3$ ) for the salient attributes than for the non-salient attributes. These ratings represent 52.3 per cent of the ratings for the salient attributes; and 45.8 per cent for the non-salient attributes.

Table 1. *Over-all mean frequencies of ratings for each point of the rating scale*

	-3	-2	-1	0	+1	+2	+3	Totals
Salient attributes	2.6	5.7	5.5	4.7	8.9	9.3	3.3	40.0
Non-salient attributes	1.9	4.9	6.3	5.8	9.6	8.0	3.5	40.0
Totals	4.5	10.6	11.8	10.5	18.5	17.3	6.8	80.0

(ii) The frequency of favourable (positive) ratings is consistently higher, both for the salient and for the non-salient attributes, than the frequency of unfavourable (negative) ratings. The percentage of favourable ratings is 53.3: of the unfavourable ones, 33.6: of the neutral ones (0), 13.1.

The statistical significance of these two trends in the data was assessed in the analysis of results.

(i) *Extreme ratings for salient and non-salient attributes*

The Wilcoxon matched-pairs signed-ranks test (Siegel, 1956) was applied in order to test the prediction that the frequency of the more extreme ratings will be higher for the salient than for the non-salient attributes. The frequencies of each subject's ratings of  $-3$ ,  $-2$ ,  $+2$  and  $+3$  for salient and non-salient attributes were taken as scores, and the significance of the differences between the pairs of scores was assessed. The prediction of higher frequency of more extreme ratings for salient attributes was confirmed at  $p$  nearly 0.01 ( $T = 21$  for  $N = 15$ , one-tailed test; there was no difference between the two scores of one of the subjects).

The same analysis was applied separately to the favourable and to the unfavourable ratings. The higher frequency of more extreme ratings for salient attributes is statistically significant for the unfavourable ratings ( $T = 14.5$  for  $N = 13$ ;  $p < 0.025$ ); the same tendency in the case of favourable ratings does not reach the level of statistical significance.

(ii) *The preponderance of favourable ratings*

The frequencies of each subject's positive and negative ratings were taken as his scores, and a Wilcoxon test applied to the differences between each pair of scores. The higher frequency of favourable ratings is statistically significant at  $p < 0.01$  ( $T = 5$  for  $N = 16$ ; two-tailed). The same analysis was then conducted separately for the scores obtained for the salient attributes and for the non-salient attributes. In both cases the higher frequency of favourable ratings is significant at  $p < 0.01$  ( $T = 9.5$  for salient attributes;  $T = 12.0$  for the non-salient ones).

#### 4. VALIDATION OF THE ASSUMPTION CONCERNING PRIORITY AND FREQUENCY

The results reported in the previous section show that the salient attributes (those which tend to appear early and to be repeated frequently in the descriptions of other people) are related to frequent use of the more extreme categories of a response scale. The assumption was made in the introductory section that priority and frequency of appearance of an attribute are indices of the degree of its importance to the subject. The confirmation of the initial prediction relating the importance of an attribute to higher frequency of extreme judgments depends therefore on the possibility of validating this assumption.

##### (i) *Method*

A separate experiment was conducted for this purpose. The first stage of the procedure was identical to the first stage of the experiment previously described: a group of 21 male subjects (students at Ruskin College) who did not take part in the first experiment described personal characteristics of young men in the set of ten photographs used in the previous experiment, and projected in the same way on a screen. An independent judge went through each subject's descriptions and reduced them to a smaller number of near-synonymous and non-overlapping clusters. Out of these, eight were again selected for each subject, four ranking highest in their combined frequency and priority, and four ranking lowest.

A list of eleven trait names was then prepared for each subject: four of these were the highest ranking attributes he had used, four the lowest ranking attributes, and three 'buffer' attributes were added, usually taken from another subject's list. These eleven attributes were then arranged in all their possible paired combinations (55 in all), and each pair written on one page of a booklet separately prepared for each subject. Using a method devised by Ross (1939), the 55 pairs were so arranged that (i) each item was paired with every other item; (ii) each item appeared equally often at the top and at the bottom of a page; (iii) the same item was never involved in adjacent pairs; and, (iv) as far as possible an item alternated from the top to the bottom of a page in its successive presentations.

Of the 55 pairs, only 16 involved a choice between one of the four highest and one of the four lowest ranking attributes; i.e., only 16 choices were directly relevant to the validation of the assumption about frequency and priority.

The second part of the experiment took place 3 weeks after the first session. Fifteen subjects were still available—some did not reappear, and some of the original descriptions could not be used as they mainly consisted of details about physical characteristics of the faces shown. The booklets were distributed to the subjects. The following instructions were given:

'The booklet in front of you contains a series of paired characteristics of people. I want you to attend carefully to each pair and to underline that item in the pair which you feel is more important in a person. Sometimes the choice will be difficult, but please make a choice for every one of the pairs given.'

The hypothesis was that if the assumption relating the subjective importance of an attribute to the priority and the frequency of its appearance was justified, then there should be a statistically significant preponderance of choices of the four highest ranking attributes as 'the more important in a person'.

##### (ii) *Results*

Table 2 sets out the relevant frequency distribution of the choices. It will be noted that none of the subjects chose less than seven high ranking attributes and that most of them chose considerably more. The mean number of such choices is 10 out of the 16 possible ones.

The Chi square test for one sample was used to assess the statistical significance



of the results, on the assumption that under a null hypothesis an equal number of subjects would choose a majority of high ranking and a majority of low ranking attributes. The observed distribution, which can be seen from Table 2, is as follows: Twelve subjects chose a majority of high ranking attributes; one subject chose an equal number from each of the two categories; two subjects chose a

Table 2. *Frequency distribution of choices of highest ranking attributes as 'more important'*

Number of Ss	Number of highest ranking attributes chosen out of 16 possible choices
3	13
2	11
4	10
3	9
1	8
2	7
—	
15	

majority of low ranking attributes. For this distribution  $\chi^2 = 7.14$ . It can therefore be concluded that the greater frequency of choice of high ranking attributes as 'more important' is statistically significant at  $p < 0.01$ .

## 5. CONCLUSIONS AND DISCUSSION

The results can be summarized as follows:

(i) Attributes which appear early and which are repeated frequently in free descriptions of other people tend to be assigned more extreme ratings than attributes which have low frequency and priority.

(ii) This is an overall finding. It conceals the fact that a significantly higher frequency of such extreme ratings was found only for the distribution of the unfavourable judgments. The favourable judgments show a tendency in the same direction, but this tendency does not reach statistical significance.

(iii) Attributes which have high ranks in terms of frequency and priority tend to be judged as 'more important in a person' than the low ranking attributes.

(iv) The ratings made by our subjects show a consistent preponderance of favourable judgments about other people.

It would be tempting to draw far-reaching and cheering conclusions from this last finding, especially as it is quantitatively more convincing than all the others. There are, however, several reservations in the way of interpreting it as a sign of universal goodwill and tolerance. First, it is just possible that our choice of photographs yielded faces which, on the whole, *were* very nice. Secondly, our subjects and the young men shown in the photographs had a good deal in common: age, nationality, probably much of the background. Thirdly, the subjects knew that their ratings would be subsequently seen by the experimenters: presumably many of them would hesitate to produce gratuitously a string of spiteful judgments about people who were unknown to them.

None of these reservations seem to hold in relation to the first finding. There are no apparent reasons which could determine a greater use of the extreme rating categories for one class of attributes rather than for another. It must not be forgotten that the *content* of these classes changed from subject to subject; an attribute which appeared as salient for one subject often appeared as non-salient for another, or did not appear at all in other lists. The only common feature of the division of attributes was the one determined in advance: in terms of the criteria of priority and frequency. Despite these variations in content, the results, though not quantitatively impressive, show a marked degree of consistency.

This consistency is more marked for the unfavourable than for the favourable judgments; but this finding is perhaps less mysterious than it appears to be. It is well in line with the initial prediction considered in the light of the data. The prediction referred to a greater use of extreme judgment categories for the salient attributes. The extreme categories were defined in advance as the two outer pairs of rating points at the favourable and the unfavourable ends of the rating scale; but, in view of the skewed distribution of judgments, the unfavourable extremes are more extreme for our group of subjects than the favourable ones. The overall median of ratings is not at the neutral point of 0, but somewhere between 0 and +1.

This skewness of extreme choices is also in line with results reported by Hovland & Sherif (1952) whose subjects tended to cluster their ratings of statements at the end of the scale far removed from their own position on the issue. This clustering tendency was less marked at the end of the scale near to the subject's own position. If one defines 'position' in our experiments in terms of the subjects' general tendency to view rather favourably the people whose characteristics they were requested to evaluate, then our finding bears a close similarity to the one reported by Hovland & Sherif.

The fourth finding concerning the assumption about priority and frequency establishes—in conjunction with the other findings—the validity of the hypothesis as initially formulated: that in the evaluation of other people's characteristics there is a tendency to use more extreme judgments for those attributes which are more important to the rater.

Two final points, related to those raised initially in the Introduction, remain to be made. First, that the findings reported here are coherent with, and extend the scope of, a general judgment phenomenon already established in related fields. Secondly, the generalizations emerging from the present study can be applied as predictions and checked without too much difficulty in a variety of 'real', out-of-laboratory situations. Knowing what is important to the rater may enable us to predict how and when he will tend to use his more extreme judgments. This information about salience is quite often available: for example, when we know someone fairly well, or when we are confronted with situations in which the aims and the functions of judgments are clearly defined.

This investigation was conducted as part of a research project supported by a grant from the Department of Scientific and Industrial Research.



## REFERENCES

- HASTORF, A. H., RICHARDSON, S. A. & DORNBUSCH, S. M. (1958). The problem of relevance in the study of person perception. In *Person Perception and Interpersonal Behaviour* (eds. Tagiuri, R. & Petrullo, L.), pp. 54-62, Stanford Univ. Press.
- HOVLAND, C. I. & SHERIF, M. (1952). Judgmental phenomena and scales of attitude measurement: item displacement in Thurstone scales. *J. abnorm. soc. Psychol.* **47**, 822-32.
- MANIS, M. (1960). The interpretation of opinion statements as a function of recipient attitude. *J. abnorm. soc. Psychol.* **60**, 340-44.
- PETTIGREW, T. F., ALLPORT, G. W. & BARNETT, E. O. (1958). Binocular resolution and perception of race in South Africa. *Brit. J. Psychol.* **49**, 265-78.
- ROSS, R. T. (1939). Optimal orders in the method of paired comparisons. *J. exp. Psychol.* **25**, 414-24.
- SHERIF, M. & HOVLAND, C. I. (1961). *Social Judgment. Assimilation and Contrast Effects in Communication and Attitude Change*. New Haven and London: Yale Univ. Press.
- SIEGEL, S. (1956). *Nonparametric Statistics for the Behavioural Sciences*. New York McGraw-Hill.
- TAJFEL, H. (1959). The anchoring effects of value in a scale of judgments. *Brit. J. Psychol.* **50**, 294-304.
- TAJFEL, H. (1962). Social perception. In *Social Psychology through Experiment* (eds. Humphrey, G. & Argyle, M.) pp. 20-54. London: Methuen.
- TAJFEL, H. & CAWASJEE, S. D. (1959). Value and the accentuation of judged differences: A confirmation. *J. abnorm. soc. Psychol.* **59**, 436-39.

*Revised manuscript received 20 March 1963*

## A Slang Vocabulary Test as an Indicator of Delinquent Association

By H. B. GIBSON

*Institute of Criminology, University of Cambridge*

The rationale of the Slater Selective Vocabulary Test was followed in devising a test which would indicate habitual delinquent association. Pilot work gave a list of slang words likely to be especially familiar to delinquent boys, and a test instrument was composed of these words buffered with items from the Mill Hill Vocabulary Scale. This instrument gives separate measures of knowledge of delinquent slang and of general vocabulary level. The test was administered to groups of normal and delinquent boys, and factor analysis of the results indicated that slang knowledge exists somewhat independently of general vocabulary knowledge. The delinquent group had a significantly greater slang knowledge relative to their vocabulary level than the normal boys. Proposals for further use of the technique are discussed.

### INTRODUCTION

It has been shown by Terman & Miles (1936) that it is possible to differentiate between the sexes by means of scores obtained on tests of general knowledge, free association and interest. Slater (1944) has developed this work in England and produced the Selective Vocabulary Test by means of which boys and girls can be allocated a percentile rank on a masculinity-femininity gradient. This test was used by Slater & Slater (1947) who demonstrated that it was possible to differentiate significantly between normal and homosexual males. The latter group had a vocabulary which was more closely akin to the female in content.

The foregoing studies have shown that vocabulary knowledge may be used as an indicator of subjects' interests and habits. We learn the meanings of words both by direct social interaction and from the indirect social stimuli of books, films, etc. For the younger and less educated person, the range of words which daily confronts him is always greater than he can comprehend; he must therefore select, and by measuring the results of his selection we can gain a knowledge of some aspects of his manner of life.

In planning the present study it was proposed to make use of a vocabulary test as a covert technique for assessing the degree to which boys are familiar with delinquent slang in comparison with their general vocabulary level. It is well known that delinquents use certain slang terms, and have special verbal usages which are not so current among the law-abiding. Much slang is ephemeral in nature, but there are certain terms (e.g. 'flash') which have an ancient and abiding significance to the delinquent. Again, ordinary words like 'straight' which have a variety of meanings, tend to be used particularly in one special sense by delinquents. It is not suggested that young boys will consciously set out to learn a 'thieves' cant', but merely that those who begin to associate with delinquent companions older than themselves, to be interested in their activities and to identify with them, will



pick up a knowledge of delinquent slang and word usage. A similar knowledge will also be acquired by the bright boy who has a precocious knowledge of words (and may even acquire it from books rather than delinquent companions), but in his case his knowledge of delinquent slang will be matched by the higher level of his general vocabulary.

## METHOD

### *Pilot work*

The first task was to compile a list of words which could be called delinquent slang. It was decided to limit the test to London slang. A pool of terms used by delinquents was collected from the police and other quarters and it was then necessary to ascertain the frequency with which the meanings of these were known to delinquent boys. Here the problem was to conceal, to some degree, just what the experimenter was after. The slang list was therefore heavily buffered with innocent words which are homonyms, and each subject, tested individually, was told that the task was to see how many different meanings he could give for each word. Thus each innocent homonym elicited a number of meanings, and so did the slang words themselves, as these are often homonyms having a respectable meaning as well (e.g. 'fiddle' means both a violin and to cheat with money). Sometimes the slang meaning was only given as an afterthought in response to the experimenter's insistence that each word had several meanings. The subjects for this pilot work were boys committed to a remand home and boys belonging to a 'tough' youth club in London.

### *Test construction*

The instrument purports to be a 'Knowledge of Words Test' and makes use both of 'delinquent' words, and of words of the Mill Hill Vocabulary Scale, Form 1 Junior, Set B (Raven, 1948). In the latter scale the subject is provided with the test words in heavy type, and for each one there is a set of six alternative meanings. His task is to underline the *one* correct meaning for each word. Alternate groups of words from the Mill Hill Scale were incorporated in the 'Knowledge of Words Test', but an amendment was made so that there were *two* correct meanings given for each word. Alternating with these Mill Hill items were groups of words which contained one item concerned with delinquent slang. The instructions are that the subject must underline *two* words which mean much the same as the word in heavy type at the top. Thus one item reads:

#### NICK

boost	harbour
clip	indent
goods yard	police station

While almost every boy will underline 'clip' as one meaning, the second choice is determined by whether the boy knows that a 'nick' is also a 'police station'. The word 'indent' is, in fact, most frequently underlined as an innocent alternative but other innocent alternatives are sometimes underlined, presumably by boys who are at a loss for a second meaning and therefore guess at random.

It was thought desirable that boys should not divine the purpose of the test, for if they became aware that it was designed to find out the extent to which they were familiar with delinquent slang, an attitude of defensiveness and concealment might arise. The scale consists of 31 groups of words, each containing seven items. Buried among these 217 items are 15 items associated with delinquency and slang. It does not seem likely, therefore, that the purpose of the test is obtrusive. In one of the groups no actual slang term is involved; the critical word is 'larceny' followed by six possible and plausible-sounding meanings. In choosing this it was thought (and subsequently confirmed) that knowledge of the correct meaning of this word among boys under 15 years of age would correlate with a knowledge of delinquent slang.

*Administration of the scale*

The scale was administered to 129 boys aged 11 to 14 years in secondary schools, and to 56 delinquent boys aged 12 to 14 years in a remand home. It is important to emphasize that the great majority of these boys had been arrested only recently, and hence they were in no sense 'institutionalized delinquents' speaking an institution slang. The manner of administration was to demonstrate on the blackboard how two words in the first group are underlined to show that they mean much the same thing as the word in capital letters at the head of it. The boys were then instructed that in all groups they were to underline two items, *guessing if they were not perfectly sure*. The latter instruction helps to ensure that all boys have an equal chance of success through random endorsements. The experimenter read aloud the whole scale so that any poor readers should not be penalized.

## RESULTS

Two scores are obtained from each protocol, (a) a measure of ordinary vocabulary level, and (b) a measure of delinquent slang.

*(a) The vocabulary score*

In marking the ordinary Mill Hill Vocabulary Scale one point is scored for each correct definition. In the altered version which was used, *two* correct alternatives are supplied in each group of six. To hit on *both* correct alternatives by chance alone, the odds are 1 in 15; to hit on *one* of the correct alternatives with two tries, the chance is 8 in 15. It would appear, then, that the scoring for the double correct definition should be very much more heavily weighted than for the single definition. In practice there was a good way of checking upon the validity of the scoring method adopted, for one of the schools provided three criterion groups of boys in the same academic stream but at three successive age levels ( $n = 82$ ). By the scoring system adopted, vocabulary level rose with age in the expected manner. The scoring system established empirically consists of allocating three points for each double definition, and one point for each single definition of the Mill Hill words.

*(b) The slang score*

A point is gained by each slang meaning indicated. It had to be established that there is a factor of slang knowledge which exists somewhat independently of general vocabulary knowledge, for otherwise the test would rest upon an unproven *a priori* assumption. To investigate this question, the data obtained from the 129 normal schoolboys, and from the first group of delinquents to be tested ( $n = 24$ ), was subjected to a statistical analysis.\* In this analysis the inter-correlations between *all* the items in the test (that is, including both Mill Hill and slang items) were computed, and from this matrix the principal components were extracted by Hotelling's method. The first three components accounted for 22.2 per cent of the variance.

In order to arrive at a simple structure, the positions of all the items were plotted with respect to their loadings on the two orthogonal axes of the first and third components ( $F_1$  and  $F_3$ ). On this graphical plot, a new axis ( $F_1'$ ) was drawn at  $45^\circ$  to the two others, and the loadings of the items on this new axis were

\* This analysis was carried out by the University of London Computer Unit.



computed. On a fresh graph,  $F_1'$  was drawn orthogonal to the axis of the second component ( $F_2$ ), and the positions of all the items were plotted with respect to their loadings on these two axes. This graphical plot is shown in Fig. 1.

The foregoing procedure is simply a method of depicting in two dimensions the structure which would be apparent in a three dimensional model constructed according to the data of the first three principal components. It so happens that the most interesting features of the present data are rendered apparent in this way. A key to the identity of the items is given in Table 1.

Table 1. *Identity of the items shown on the factorial plot in Fig. 1*

Mill Hill items	Delinquent items
2. Rest*	1. Drag = puff of cigarette
4. Afraid*	3. Grass = tell tales
6. Blaze	5. Flash = tough
8. Squabble	7. Spade = coloured man
10. Shrivell	9. Nick = police station
12. Provide	11. Chiv = razor
14. Schooner	13. Fiddle = cheat with money
16. Courteous	15. Hop the wag = skip from school
18. Thrive	17. Snout = cigarette
20. Dwindle	19. Screw = rob a house
22. Whim	21. Straight = honest
24. Bombastic	23. Lift = take from a shop
26. Envisage	25. Fly = crafty
28. Glower	27. Larceny = robbery
	29. Receiver = pawnbroker (or fence)
	29a Membership of delinquent group

\* These items were not included in the analysis as they were endorsed by nearly 100 per cent of the subjects.

It may be seen by inspection of Fig. 1 that there are two recognizable clusters, Mill Hill items (even numbers) and delinquent items (odd numbers). This confirms the hypothesis that slang knowledge exists somewhat independently of general vocabulary knowledge. Membership of the delinquent group (29a) is positioned in the appropriate cluster.

As the scatter of slang items on the factor plot is rather diffuse, it was decided not to use any system of weightings in forming the slang scale. Item 1 was omitted from the final scale, as the slang meaning was known by 91 per cent of the population, and therefore adjudged to be too 'easy'.

In the allocation of appropriate total slang scores, a correction must be made for general vocabulary level. The regression of the raw slang score on Mill Hill vocabulary score was therefore calculated ( $b_{y,x} = 0.166$ ) and the effect of the latter was partialled out. The advantage of this procedure is that, although both general vocabulary and slang vocabulary increase with age, what is being measured is the *relative* amount of slang known by the individual.

Table 2 shows the data on the slang scores (adjusted for regression) of the normal and delinquent groups divided into four age levels.

It may be seen from Table 2 that there is little meaningful relationship between

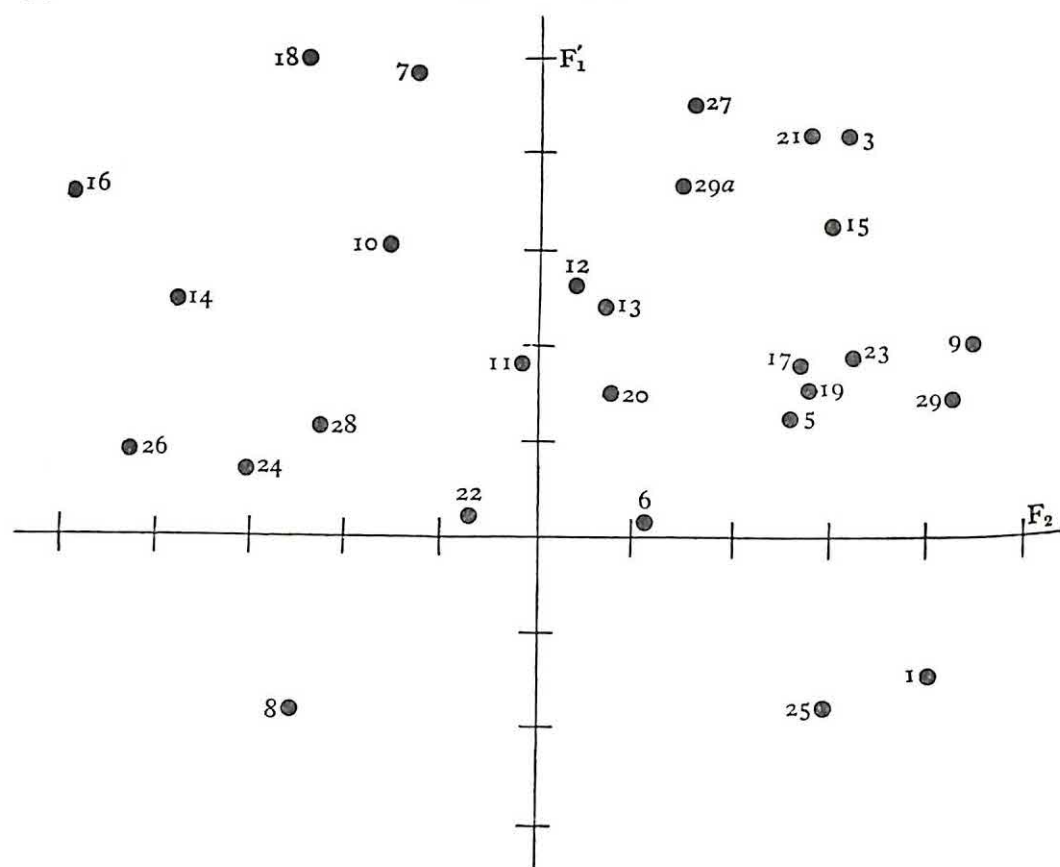


Fig. 1. Loadings on two factors. Mill Hill items (even) and delinquent items (odd).

age and the adjusted slang scores. The 11-year-old group has the lowest mean score, but in neither the normal nor the delinquent group does the mean score rise consistently with age. At all age levels the delinquent group had a significantly higher score.

Table 2. *Slang scores (adjusted for regression) of the normal and delinquent boys at four age levels*

	11 years			12 years			13 years			14 years		
	<i>n</i>	Mean	S.D.	<i>n</i>	Mean	S.D.	<i>n</i>	Mean	S.D.	<i>n</i>	Mean	S.D.
Normals	36	6.79	3.07	52	7.34	2.11	33	7.42	2.77	8	6.83	1.52
Delinquents				12	9.58	1.64	24	8.89	1.62	20	9.25	1.28
Significance of difference between means				$p < 0.001$			$p < 0.05$			$p < 0.001$		

## DISCUSSION

The results of this study show that it is practicable to measure boys' knowledge of slang and to relate it to their general vocabulary level. The index of slang knowledge thus obtained has been shown to discriminate significantly between normal schoolboys and boys arrested for delinquency. The test which has been developed



is adequate enough to be usable and may prove a valuable tool for identifying potential or undetected delinquents.

In the construction of the Slater Selective Vocabulary Test the percentage of boys' and girls' endorsement was calculated separately for each item, but in the present study the rationale of test construction was somewhat different. Although the delinquent group could be shown to endorse the critical items more frequently than the normal groups, this was not the justification for grouping these items into a slang scale. The items were chosen in the first place because a pilot group of delinquents and 'tough' boys had shown knowledge of their slang meaning, but the justification for including such words in a scale lies in the fact that they clustered together separately from the Mill Hill words when the data were factor analysed. It might have proved that these slang words were equally well known to the normal schoolboys, and the contrary was established only when the distributions of the corrected slang scores for the two groups were compared in the final analysis.

The limited nature of these results must be recognized. Only 14 of the critical words were used in the final scale, and it is to be hoped that further work of this kind will produce a longer list, thus adding to the discriminative power of the test. Again, in the present study the Mill Hill Vocabulary Scale was not used in its full and proper form, since it was first necessary to see whether a selection of Form 1 Mill Hill words, used as buffers, would form a cluster separate from the slang words on factor analysis. Such a selection does not give as reliable a measure of general vocabulary level as the proper scale. A more precise measurement of vocabulary level is needed in order to correct the raw scores of the slang scale if this scale is to discriminate accurately between *individuals*. However, it would be perfectly feasible to administer Form 2 of the Mill Hill Scale in the orthodox manner in addition to the combined scale.

In spite of the limitations of this study, publication of these results at the present stage has been deemed advisable so that the instrument shall be available to other research workers and in order that research in this field may be stimulated.

I am grateful to the Children's Department of the L.C.C. for their kind co-operation in the provision of facilities for this study.

#### REFERENCES

- RAVEN, J. C. (1948). *The Mill Hill Vocabulary Scale*. London: H. K. Lewis.  
SLATER, P. (1944). *The Selective Vocabulary Test*. London: G. G. Harrap.  
SLATER, E. & SLATER, P. (1947). A study in the assessment of homosexual traits. *Brit. J. med. Psychol.* 21, 61-74.  
TERMAN, L. M. & MILES, C. C. (1936). *Sex and Personality*. New York: McGraw Hill.

*Manuscript received 18 March 1963*

*Revised manuscript received 25 March 1963*

## Cultural Comparisons of Boys' Occupational Aspirations

By WALLACE E. LAMBERT *McGill University, Montreal*

AND

OTTO KLINEBERG *Université de Paris*

The occupational aspirations of samples of 6-, 10- and 14-year-old boys from eleven cultural settings were compared. Certain aspirations were popularly mentioned in most cultures represented although several exceptions to this general trend were noted. Each boy's aspirations was compared to his father's actual occupation, and 'filial-aspiration' indices were calculated, reflecting a tendency to aspire below, above or at the same social-class level as that represented by the father's occupation. Reliable differences in the distribution of these indices were noted for culture and age variations. Filial-aspiration indices were found to be significantly correlated with an independent measure of cultural attitudes toward achievement. The results are interpreted as support for the hypothesis that cultures vary in the freedom given children to modify established levels of the family's social standing and that filial-aspiration indices reflect these cultural differences.

Recent advances in the study of motivation stem in large part from the use of a new approach, one which examines the manifestations of a particular motive in various behavioural settings and traces its development to various complex antecedent sources of influence. The work of McClelland, Atkinson, Clark & Lowell (1953) on the achievement motive is an outstanding example of this approach. In this study, they compared various situational conditions with regard to power of evoking the achievement motive. Later work (McClelland, 1955*a, b*) extended the scope to societal and cultural (including philosophical and religious) variables which affect the amount of achievement motive likely to be found among members of various religious, social class, or cultural groups. The results of these investigations strongly support the hypothesis that achievement orientations, conceptualized as stable personality dispositions, are developed through particular socialization influences which, in turn, are affected by cultural and philosophical values. The stability of this disposition has been established by an important recent study by Kagan & Moss (1959), who showed that achievement strivings noted at ages 6 to 10 are highly related to similar tendencies measured in adulthood.

In the same spirit, Rosen (1959) has examined the interrelations of achievement motives, cultural values toward achievement, and the educational and occupational

\* The information used in the present study was gathered by a number of social scientists in various national settings while they were interviewing children as part of a larger investigation being conducted by the authors. The whole programme of research was supported by subventions from UNESCO, the Research Institute for the Study of Man, and Canada Council. The following were responsible for the collection of data in the various centres: Dr Rafia Semin, Istanbul; Dr T. Prothro, Beirut; Dr W. E. Lambert, Montreal; Dr S. N. Eisenstadt, Jerusalem; Dr Hazel D. Rosenthal, Johannesburg; Dr R. Ribeiro, Recife; Dr R. Alpert, Boston; Dr Rosette Avigdor-Coryell, Paris; Dr K. S. Sodhi, Berlin; and Dr K. Sato, Kyoto.

We are particularly grateful to Dr G. A. Ferguson and Grace Yeni-Komshian of McGill for advice and assistance in the analysis and interpretations of the materials presented here.



aspirations parents hold for their children. In view of the socio-cultural histories of various ethnic groups, Rosen believed that cultural backgrounds should play a strong rôle in determining the extent of achievement striving. Comparing second-generation Americans of different racial, religious and ethnic backgrounds, Rosen's findings do not make clear whether cultural or religious backgrounds play the more important rôle in determining achievement motivation. But some very suggestive relations do exist between the stresses placed on achievement and independence training by mothers of different ethnic-religious backgrounds and the amounts of achievement motivation shown by their children. The data which follow, based on the behaviour of members of cultural groups 'at home' rather than as second-generation subgroups within the American cultural setting, throw further light on the nature of achievement motivation and its development. The results of the present study also lend support to McClelland's most recent work (1961) on the relation of achievement motivation and successful entrepreneurship in modern nations.

The present study capitalized on an opportunity to make cultural comparisons of motivation. As an adjunct to a larger study being conducted in various nations, boys in eleven cultural settings were asked two questions, the responses to which provided a measure of a particular aspect of motivation. One question dealt with the boys' occupational or vocational aspirations, or what each boy wanted to do for his livelihood when he would be an adult, and the other determined what each boy's father actually did for his livelihood. The boys' aspirations and their fathers' occupations were compared in terms of social class similarities. A son might aspire to the same occupation as his father's or one judged as socio-economically equivalent for that culture, one of a higher social class than that of the father, or one lower in class than the father's. The occupational comparisons, to be referred to as 'filial-aspiration' indices, were made by social scientists in each nation who were aware of the position of various occupations in that nation's economic class structure.

The purposes of the present study, then, were: (a) to compare the occupational aspirations of boys from a wide range of cultural settings, (b) to relate boys' aspirations to the social status of their fathers' actual occupations and to examine age and cultural variations in this son-father comparison, and (c) to relate variations in these filial-aspiration indices to cultural measures of achievement motivation.

## METHODS

For the purposes of the larger study mentioned above, a varied but non-representative sample of eleven cultures of the world was chosen after consultation with a panel of eminent American social scientists. (It was unfortunate that none of the social scientists within the Communist sphere of influence who were asked to participate were willing to do so.) Boys included in any cultural sample had to live in a large city, have at least normal intelligence, and to attend publically financed schools. Three age samples of 50 boys each were chosen. Thus, 50 6-year-old, 50 10-year-old and 50 14-year-old boys were chosen from public schools in major cities in each of eleven cultural settings. Half of the boys at each age level came from what was considered as lower or 'working'-class families for the particular cultural group involved, and half came from middle-class families. The directors of the national studies collaborated with local school administrators to decide which children came from middle or lower socio-economic class homes. The final samples were considered by the directors of



each national study to be representative of urban children with normal intelligence living in the particular cultural settings involved.

The fathers' actual occupations were reported in as much detail as possible by the children themselves in interviews, and their statements were checked against school records of family backgrounds if they seemed suspicious in any way. Somewhat later in the interview, each boy was asked what occupation he desired for himself when he would be an adult. The research teams in each nation compared the son's aspirations and the father's occupation and assigned a rating of 2 if the son aspired to an occupation which would be regarded as having the same general status significance as that of the father by natives of the country in question, or a rating of 1 if the aspiration fell at a status level recognizably lower than that of the father, or a 3 if the aspiration was at a status level recognizably above the father's occupation. The raters drew on their own experience as social scientists and as residents of their country to determine if a child's aspiration would be typically recognized by others in the country as essentially the same status level as the occupation of the father or whether it would be generally considered as higher or lower in status. This within-family occupational comparison will be referred to as a 'filial-aspiration' index.

## RESULTS AND DISCUSSION

### *Cultural comparisons of boys' occupational aspirations*

The more popularly given aspirations of boys from various cultural backgrounds are presented in Table 1. Only those aspirations which were given by at least two children from an age group of any particular culture are listed.

Although the samples are too small for normative comparisons, still several general statements can be made about trends apparent in these data. First, when one disregards age and class differences, there is a general similarity of occupational aspirations among cultural samples. That is, engineering, medicine and mechanical occupations are popular aspirations in most of the cultures represented. Cultural exceptions to these generally popular occupations are seen in the recurrent choices of the priesthood among the two younger age groups of French-Canadian boys, of teaching among the Bantu boys, of manual type occupations among German boys, and of employee or sales positions or baseball careers among the Japanese boys.

Secondly, distinctive social class variations do not, in general, show themselves in occupational choices. The lower- and middle-class boys appear to have very similar aspirations in most cultures represented. For example, the choice of engineering and medicine is generally as popular for lower- and for middle-class boys. Exceptions to this trend are noticeable in the German and French older samples where the lower-class children more typically aspire to lower-class positions while the middle-class children mention middle-class aspirations.

Thirdly, there are no outstanding age variations in occupational aspirations. Although the younger children, comparatively, are apparently attracted to the more glamorous and adventurous occupations (such as soldier, pilot, fireman, policeman and, particularly, truck driver) still the more popular choices of the older boys for any culture are also apparent among the younger boys of that culture. Thus there is the suggestion of a progressive age trend from the more adventurous and glamorous occupations toward the more sober and mature occupational aspirations.

The interesting fact is the similarities of occupational aspirations noted among children of various ages, from different socio-economic backgrounds and from cultures that vary in degree of economic development and standard of living. The wide-spread similarity of aspirations suggests the existence of a universal view



Table 1. Popular occupational aspirations for boys from eleven cultural settings

Boys' age:	6 years			10 years			14 years			
	Social class:			Social class:			Social class:			
	Low	Middle		Low	Middle		Low	Middle		
TURKEY	soldier	7	engineer	8	11	engineer	9	11	engineer	11
	doctor	4	navy	3	8	architect	4	3	doctor	4
	engineer	4	pilot	3	3	police	3	2	police	2
	employee	2	architect	2		doctor	2			
LEBANON			doctor	2		pilot	2			
	soldier	5	doctor	11	5	doctor	10	12	engineer	10
	doctor	4	pilot	4	4	engineer	9	5	doctor	6
	teacher	3	merchant	3	3	soldier	2	3	pilot	3
FRENCH CANADA			engineer	2	2	merchant	2	2	soldier	2
						painter	2		teacher	2
						(house)	2			
	construction	9	police	5	4	priest	6	5	engineer	10
ISRAEL	priest	5	doctor	3	2	doctor	3	3	electro-nics	3
	soldier	2	mechanic	3	2	farmer	3	2	mechanic	2
			priest	3	2	police	3	2	printer	2
					2	reporter	3	2	teacher	2
						accountnt.	2			
						pilot	2			
						chemist	2			
	engineer	4	soldier	6	4	pilot	5	4	engineer	8
	pilot	4	doctor	4	2	driver	3	4	doctor	2
	driver	3	engineer	3	2	soldier	2	3	lawyer	2
	police	2	driver	2		sailor	2	3	farmer	2
	mechanic	2	pilot	2		mechanic	2	2		
	carpenter	2				engineer	2	2	teacher	2
								2	farmer	2

Table 1.—continued.

Boys' age:	6 years			10 years			14 years			
	Low		Middle	Low		Middle	Low		Middle	
Social class:	teacher	12	teacher	6	teacher	6	teacher	7	clerk	6
	doctor	3	driver	6	doctor	4	doctor	4	teacher	6
	driver	2	doctor	5	driver	3	clerk	3	doctor	3
			clerk	4	carpenter	2	clerk	3	driver	3
			builder	2			policeman	3	lawyer	2
BANTU										
	policeman	5	policeman	5	pilot	4	engineer	5	engineer	4
	ballplayer	2	fireman	4	athlete	2	doctor	4	electric-	2
	doctor	2	doctor	3	doctor	2	police	4	ian	2
	engineer	2	navy	2	marine	2	ballplayer	3	mechanic	2
U.S.A.	milkman	2	teacher	2	priest	2	priest	2	pilot	2
	sailor	2					scientist	2	teacher	2
	soldier	5	doctor	8	doctor	5	engineer	10	mechanic	9
	chauffeur	3	engineer	4	engineer	4	doctor	8	engineer	8
BRAZIL	taylor	2			mechanic	2			typesetter	2
	mechanic	2			chauffeur	2				
	doctor	2								
	dentist	2								
ENGLISH CANADA	policeman	4	doctor	6	engineer	6	engineer	3	engineer	8
	sailor	2	engineer	4	pilot	4	soldier	2	doctor	4
	train en-		policeman	4	scientist	2	lawyer	2	lawyer	3
	gineer	2	pilot	2			pilot	2		



FRANCE	None of 2 each	fireman	3	mechanic	3	engineer	10	mechanic	6	doctor	4		
		barber	2			doctor	7	engineer	2	draughtsman	2		
		doctor	2			pilot	2	electronics	2	machinist	2		
		policeman	2					locksmith	2				
	GERMANY	actor fireman mason					metal worker	2					
							house painter	2					
							toolmaker	2					
			2	policeman	4	butcher	2	train con-		baker	4	electron-	5
			2	driver	3	engineer	2	ductor	4	mechanic	4	ics	4
			2	pilot	2	driver	2	policeman	3	civ. serv.	3	engineer	4
JAPAN	driver engineer merchant clerk metal worker pilot	soldier	2	worker	2	engineer	2	construction	2	teacher	3		
								engineer	2	architect	2		
								merchant	2	doctor	2		
								telephone	2	sailor	2		
		5	baseball	4	clerk	6	baseball	6	clerk	6	clerk	7	
		4	clerk	3	baseball	4	driver	4	merchant	5	sales	3	
		3	sales man-		sales	3	engineer	2	engineer	3	scholar	3	
		3	ager	3	manager	3			iron worker	2	engineer	2	
		2	driver	3	driver	2							
		2	police	3	police	2							
	salesman	3											

Table 2. *Filial-aspiration indices for boys from eleven cultural settings*

Culture	6 years		10 years		14 years		Culture means		Achievement scores†	
	$\bar{X}$	N	$\bar{X}$	N	$\bar{X}$	N	$\bar{X}$	N	Means	Standard scores
Turkey	2.44	45	2.75	48	2.81	48	2.67	141	3.62	2.16
Lebanon	2.45	42	2.46	56	2.68	59	2.54	157	2.71	0.95
French Canada	2.38	42	2.58	38	2.63	35	2.52	115	3.10	1.46
Israel	2.24	45	2.46	39	2.78	46	2.50	130	2.33	0.44
South Africa (Bantu)	2.42	50	2.58	50	2.42	45	2.48	145	None available	
Brazil	2.24	45	2.59	46	2.43	46	2.42	137	None available	
United States	2.26	43	2.49	43	2.50	40	2.41	126	2.24	0.32
English Canada	2.11	36	2.34	44	2.29	35	2.25	115	2.29	0.39
France	2.03	34	2.43	42	2.17	42	2.22	118	2.38	0.51
Germany	2.02	44	2.13	45	2.38	60	2.20	149	2.14	0.19
Japan	1.83	47	2.07	45	2.48	44	2.12	136	1.29	-0.95
Age means	2.23	473	2.45	496	2.51	500				

Minimum mean age difference\*:  $p < 0.01 = 0.08$ ,  $p < 0.05 = 0.07$ .

Minimum mean cultural difference\*:  $p < 0.01 = 0.16$ ,  $p < 0.05 = 0.15$ .

\* These minimum differences were determined following the techniques of Sheffe as presented in Edwards (1960, p. 154).

† Taken from McClelland (1961, pp. 461-463). McClelland had determined a score for Brazil but it was decided not to use it for present purposes because McClelland noted that it was 'not used in the main study, scored separately and open to coding bias' (p. 463). Three scores were given for Canada, one labelled 'Canada' which was used for the English-Canadian entry since it likely refers to readers used by Montreal's Protestant School Board and its system of schools from which the English-Canadian children were chosen. Two other scores were given, an 'English Catholic' and 'French Catholic'. Since training in English is started early (grade 4) in the Montreal French schools, and since school books must be approved by the Catholic School Board, it was decided to average the scores for the French and English Catholic readers since both are likely used in the training of the French-Canadian children. It was not possible to do the same for the English-Canadian sample who also study French in the early grades, because McClelland had no score available for French Protestant readers. It is unlikely that Catholic books would be used in the Protestant schools of Montreal.



among children, regardless of their backgrounds, of what occupational rôles are of interest.

*Cultural comparisons of filial-aspiration indices*

Cultural variations in mean filial-aspiration indices are presented in Table 2, where cultures are ranked from high to low in this regard. For this analysis it was decided to combine the data for lower- and middle-class boys for several reasons. First, the filial-aspiration indices actually collected in the different cultural settings only permitted three degrees of variation: the son's occupational aspiration was categorized as being above, below or at the same social-class level as was his father's occupation. More gradations of comparisons could, of course, be used in determining filial-aspiration indices, but in the present study it was not feasible to obtain more refined occupational comparisons. Consequently, the middle-class boys have less opportunity of aspiring above their father's occupational level than do lower-class boys, and lower-class boys are similarly limited at the other extreme in that

Table 3. *Variance analysis for culture and age variations in filial-aspiration indices\**

Source of variation	Sum of squares	df.	Mean square	F	Significance
Culture	40.85	10	4.085	11.1216	<0.01
Age	21.882	2	10.941	29.7876	<0.01
Interaction	13.473	20	0.6736	1.8339	<0.05
Within	527.504	1,436	0.3673		
Total	603.71	1,468			

\* Because the N's varied from cell to cell, the methods for equating cell entries suggested by Ferguson (1959, p. 259 f) were carried out. The N's varied primarily because of children who gave no aspirations, i.e., they didn't know what they wanted to be when adults.

they have less opportunity of aspiring below their fathers' occupational level. Social class comparisons would have been affected by these limitations and were accordingly not made. Instead, attention is directed here to cultural and age variations based on samples of approximately equal numbers of lower- and middle-class boys at the three different age levels.

Table 3 presents the results of an  $11 \times 3$  analysis of variance of the distribution of filial-aspiration indices. It will be noted that there are reliable differences among cultural and age groups of boys and that the age and cultural variations interact. The minimum cultural differences needed for significance (presented in Table 2) indicate that there are two clearly distinguished cultural clusters. Those having high filial-aspiration indices are: Turkey, Lebanon, French Canada, Israel, the Bantu sample, Brazil and United States, in that order. Turkey is reliably higher in this respect than all other cultural samples except Lebanon. The cultures with comparatively low filial-aspiration indices are: English Canada, France, Germany and Japan, in that order.

With respect to age differences, it will be noted that there is a marked increase in the magnitude of the indices from 6 to 10 years of age whereas the 10- to 14-year increase does not quite reach significance levels. Thus, the 6-year-old boys have lower filial-aspiration indices than do the older boys. The interaction effect noted in Table 3 is due to age trend differences from culture to culture. In six of the

cultural samples (Turkey, French Canada, Israel, Germany and Japan) there are step-wise increases from year to year in the indices, while in four samples (Bantu, Brazil, English Canada and France) the highest indices are noted at age 10 with a decrease from 10 to 14 years.

The filial-aspiration index reflects the magnitude of discrepancy between a father's occupational achievement and his son's aspiration. In bringing these two elements into a ratio, it is assumed that the son sets his aspirational goals partly in terms of his father's actual occupation, in some cases lining up his goals with his father's achievement level, and in others departing from it by orienting above or below this parental reference point. Furthermore, we are placing emphasis on the average filial-aspiration index of groups of children from different cultural settings. It is assumed that these mean indices reflect variations in cultural orientations toward the next generation's getting ahead in relation to the present generation, as one extreme of the dimension, or staying close to the parental pattern and perpetuating the actual social class standing of the family. In other terms, it is presumed that the index reflects a cultural attitude toward the maintenance or rejection of established levels of family class status. If this attitude has culture-wide acceptance, one would expect that parents hold the attitude and try to pass it on to their children. The attitude, however, could be entertained only by certain age-groups of children regardless of the values of their parents. In this respect, the index could reflect a feeling of disappointment by the younger generation toward the cultural achievements of the older generation, a sort of looking back one generation with chagrin if not with anger.

It may clarify these notions if we try to apply them in an interpretation of two trends apparent in the data of Table 2. The French samples show a reliable decrease in filial-aspiration indices for the 10- to 14-year-old groups ( $t = 2.06$ ,  $p < 0.05$ ) while Japanese, German and Israeli boys show reliable increases in scores from 10 to 14 years ( $t = 3.90$ ;  $t = 2.50$ ; and  $t = 2.39$  respectively). These trends might mean that transmitters of the French culture curtail any inappropriate aspirational ventures by the time a boy reaches his teens whereas Japanese and German socializers may lose some type of control over their children's aspirations after the early years; the teen-age boys in the Japanese and German samples seem no longer to keep their aspirations in line with their fathers' occupations, as was apparently the case at ages 6 and 10. In Israel, there apparently is social support for high filial-aspirations which gets progressively stronger at each age level. We have no independent evidence available to determine degrees of rebelliousness of children in cultural settings like Japan or Germany. This matter calls for further research specifically designed to test the notion that socializers in various cultures differ in their attitudes toward their children's desires to get ahead.

#### *The relation of filial aspirations to achievement values*

McClelland (1961) has recently collected samples of children's readers from various countries and has analysed their content for achievement themes. The readers were those in popular use in 1950 in each country. This time period seems especially appropriate for our present purposes since all three age groups would



probably have used these readers. McClelland's sample of nations included nine of the eleven used in our own study. The means and standard scores for *n* Achievement (derived from the analysis of the children's readers) for the nine nations are presented in the right-hand section of Table 2.

The datum of interest here is the correlation between mean cultural filial-aspiration indices and McClelland's *n* Achievement score:  $\rho = 0.87$ , significant at beyond the 0.01 confidence level. This correlation reflects a large amount of common variance between children's desires to get ahead and cultural attitudes toward achievement, as these factors are measured here. The correlation indicates that the more traditionally settled nations (Japan, Germany and France) are less likely to generate this desire to get ahead whereas those cultures changing from an old to a new order, such as those in the Middle East, are more likely to stimulate in their children a desire to get ahead.

Assuming that the popularity of the readers in any cultural setting is a reflection of the culture-wide values which educators have chosen to stress, then we may consider this correlation as supporting evidence for cultural variations in attitudes towards the maintenance or rejection of established levels of family class status. The finding suggests that socializers in various cultural settings differ in the degree to which they encourage their children to get ahead in relation to the levels of achievement attained by the current adult group. The finding also suggests that achievement orientations may depend upon cultural variations in the freedom afforded children to modify established levels of the family's social standing.

#### REFERENCES

- EDWARDS, A. L. (1960). *Experimental Design in Psychological Research*. New York: Rinehart.
- FERGUSON, G. A. (1959). *Statistical Analysis in Psychology and Education*. New York: McGraw-Hill.
- KAGAN, J. & MOSS, H. A. (1959). Stability and validity of achievement fantasy. *J. abnorm. soc. Psychol.* **58**, 357-364.
- MCCLELLAND, D. C. (1955a). The psychology of mental content reconsidered. *Psychol. Rev.* **62**, 297-303.
- MCCLELLAND, D. C. (Ed.) (1955b). *Studies in Motivation*. New York: Appleton-Century-Crofts.
- MCCLELLAND, D. C. (1961). *The Achieving Society*. Princeton, N.J.: D. Van Nostrand Company, Inc.
- MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A. & LOWELL, E. L. (1953). *The Achievement Motive*. New York: Appleton-Century-Crofts.
- ROSEN, B. C. (1959). Race, ethnicity, and the achievement syndrome. *Amer. sociol. Rev.* **24**, 47-60.

*Revised manuscript received 14 March 1963*

## The Effect of the Ethnic Grouping of the Experimenter upon Children's Responses to Tests of an Ethnic Nature

By GRAHAM M. VAUGHAN

*Victoria University of Wellington, New Zealand*

The responses of 240 Maori and pakeha (White) children to tests of ethnic awareness and attitudes were examined in relation to the ethnic grouping of *E*. Half of the children of each ethnic group were tested by a Maori *E* and half by a pakeha *E*. With one exception, the *E* variable did not affect children's performance on awareness tests. On attitude tests, however, a defensive reaction by older Maori children, who favoured pakeha figures more frequently when *E* was pakeha rather than Maori, was noted. No such shift in response was observed in pakeha children. The results indicate that in ethnic attitude research, particularly when subjects used are drawn from a minority group, the provision of own-race interviewers should be an automatic control.

### INTRODUCTION

In interview situations which involve psychological testing, the adequacy of the interviewer is often judged in terms of his administrative skill. An instance in which this criterion may be inadequate is when an attempt is being made to measure ethnic attitudes, since it is possible that *Ss'* responses may be determined in part by the interviewer's race. The scant attention paid to the effects of this variable in attitude research, particularly in studies where *S* and *E* differ by ethnic group, has been described as a major design fault (Dreger & Miller, 1960). The available American evidence (Heine, 1950; Trent, 1954; Rankin & Campbell, 1955) suggests that this conclusion is warranted.

The present study examined the effects of varying the ethnic grouping of *E* upon the performance of Maori and pakeha (white) New Zealand schoolchildren on a number of tests of ethnic awareness and attitudes. Research has already indicated the existence of unfavourable ethnic attitudes in New Zealand, among pakeha students (Thompson, 1959; Vaughan & Quartermain, 1961; Vaughan, 1962), pakeha children (Vaughan & Thompson, 1961; Vaughan, 1964*b*), and Maori children (Vaughan, 1964*b*). All of these studies employed own-race *Es*. In the present study, other-race *Es* were introduced and *Ss'* responses were to be compared with those of groups tested by own-race *Es*.

### METHOD

#### *Subjects*

The *Ss* were 240 children, 80 at each of the CA levels, 7, 9 and 12, chosen from the Wellington and Hutt urban areas, whose combined population was 237,000 pakehas and 3,500 Maoris in the year 1960. The use of these age levels was suggested by previous findings of significant changes in ethnic awareness and attitudes in the range CA 6-12 (Vaughan, 1963, 1964*a*, 1964*b*). There were an equal number of Maori and pakeha *Ss* at each of the three levels, and groups were divided evenly by sex. All *Ss* were enrolled at primary schools. Pakeha *Ss* were selected at random from class rolls, while all Maori *Ss* at the schools approached were included.



*Experimenters*

Two *Es* were used, one a Maori and the other a pakeha. The author acted as pakeha *E* and a psychology student recommended by the Department of Maori Affairs acted as Maori *E*. The latter was darkskinned and of full Maori parentage. Both *Es* had had previous acquaintance with members of the ethnic out-group and were practised in establishing rapport in an interview situation. The *Es* were of similar physical stature and hair colour, and both were dressed alike during testing. Matters of procedure were rehearsed to ensure comparable conditions of administration.

Half of the *Ss* of both ethnic groups at each CA level were tested by the Maori *E* and half by the pakeha *E*.

*Materials*

Seven tests of ethnic awareness and three tests of ethnic attitudes, constructed of doll and picture materials, were used.

The awareness tests included: (1) *Picture Identification*: Which of two figures looks like *S*? (2) *Doll Identification*: As for (1). (3) *Picture Discrimination*: Which one of three figures is different? (4) *Doll Discrimination*: As for (3). (5) *Doll Assembly*: Make up two dolls from jigsaw pieces. (6) *Doll Classification (A)*: What kind of doll (brown-skinned, in traditional Maori costume) is this? (7) *Doll Classification (B)*: As for (6) but in pakeha clothing.

The attitude tests included: (1) *Stereotypes*: Which figure, of each of six pairs, is more 'lazy', 'clever', etc.? (2) *Picture Preference*: Choose a playmate one from each of two pairs. (3) *Doll Preference*: Choose one from a pair to take home.

Further details concerning the structure and implications of these tests are available (Vaughan, 1963, 1964a).

*Procedure*

The tests were administered in an order which distributed the attitudinal measures through the test series. *Ss* were tested individually with an average time of 15 minutes.

## RESULTS AND DISCUSSION

*Ethnic awareness*

A broad outline, only, of results in this section are given.

The mean proportion of *Maori Ss* making correct responses to each awareness test with a Maori *E* (and with a pakeha *E*) were: Picture Identification, 0.68 (0.57); Doll Identification, 0.73 (0.72); Picture Discrimination, 0.64 (0.73); Doll Discrimination, 0.93 (0.98); Doll Assembly, 0.88 (0.87); Doll Classification (A), 0.92 (0.85); Doll Classification (B), 0.83 (0.45). Of these means, a difference by the *E* variable is significant only on the Doll Classification Test (B), where analysis for the three CA groups shows  $\chi^2 = 17.21$  (d.f. = 2,  $p < 0.001$ ). This solitary finding may stem from shyness in Maori *Ss* when their task is to verbalize the concept 'Maori' for a pakeha *E*. (Such verbalization is required by the Classification Tests only.) Maximal cues, i.e. traditional Maori costume, in the Doll Classification Test (A) do not produce the same effect, where failure to respond might suggest ignorance in *S*.

Comparable figures for *pakeha Ss* with a Maori *E* (and with a pakeha *E*) were: Picture Identification, 0.98 (0.95); Doll Identification, 0.85 (0.93); Picture Discrimination, 0.71 (0.71); Doll Discrimination, 0.97 (0.97); Doll Assembly, 0.95 (0.93); Doll Classification (A), 0.93 (0.90); Doll Classification (B), 0.72 (0.82). None of these differences by the ethnic grouping of *E* are significant.

Generally, these findings do not indicate a significant relationship between the *E* variable and *Ss*' performance on awareness tests. Other trends not relevant to

this context, e.g. effects of age and ethnic grouping of *S*, are discussed elsewhere (Vaughan, 1963, 1964a).

### *Ethnic attitudes*

The proportions of *Maori Ss* at each CA level favouring own-race figures on each of the attitude tests, and the *E* variable, are shown in Table 1.

Table 1. *Proportions of Maori children favouring own-race figures on three attitude tests in relation to ethnic grouping of experimenter*

Test	Ethnic grouping of <i>E</i>	Age level			Mean
		7	9	12	
1. Stereotypes (Diff.: $K_D$ )	Maori	0.33	0.33	0.55	0.40
	pakeha	0.25 (5)	0.39 (4)	0.37 (7)*	0.34 ( $D = 0.183$ , $\chi^2 = 4.02$ , d.f. = 2)*
2. Picture Preference (Diff.: $\chi^2$ )	Maori	0.20	0.15	0.60	0.32
	pakeha	0.20 (—)	0.30 (—)	0.15 (15.88)‡	0.22 (1.07)
3. Doll Preference (Diff.: $\chi^2$ )	Maori	0.40	0.60	0.80	0.60
	pakeha	0.35 (—)	0.70 (—)	0.45 (3.84)†	0.50 (0.84)

\*  $p < 0.20$

†  $p < 0.05$

‡  $p < 0.001$

It will be observed in Tables 1 and 2 that the Chi Square test is used in analysing the nominal data of the two Preference Tests. The structure of the Stereotypes Test, however, permits the ordering of *Ss* in terms of the number of occasions (six through zero) that own-race figures are favoured. In this instance, maximum information is yielded by applying the Kolmogorov-Smirnov test.

These results indicate that *Maori Ss* favour own-race figures less frequently on each test when *E* is pakeha. On overall means this trend is not significant, although marginally so ( $p < 0.20$ ) for the Stereotypes Test. But by CA level the effects of the *E* variable are slight at CA 7 and 9, yet marked at CA 12. At the latter level *Maori Ss* respond in a way which is influenced by *Es* ethnic grouping: a surface attitude favouring other-race figures is promoted by employing an other-race tester. This suggests that the older *Maori* child uses a defence of a protective-withdrawal kind in an initial contact with a member of the ethnic out-group. Perhaps repeated equal-status contacts with a pakeha would reveal latent attitudes, but evidence pertinent to this issue is lacking.

The proportions of *pakeha Ss* at each CA level favouring own-race figures on each of the attitude tests, and the *E* variable, are shown in Table 2.

Inspection of Table 2 shows that *pakeha Ss* favour own-race figures more frequently on the Stereotypes and Picture Preference Tests, when *E* is pakeha, with a reverse trend on the Doll Preference Test. These effects, however, are not



significant. The same holds for an analysis by CA level. This finding could indicate that pakeha Ss perceive themselves as members of a privileged ethnic majority (cf. Morland's (1958) observation concerning American Whites), and that consequently a defence of the kind associated with older Maori Ss would serve no useful function.

Table 2. *Proportions of pakeha children favouring own-race figures on three attitude tests in relation to ethnic grouping of experimenter*

Test	Ethnic grouping of <i>E</i>	Age level			Mean
		7	9	12	
1. Stereotypes (Diff.: $K_D$ )	Maori	0.82	0.64	0.57	0.68
	pakeha	0.77	0.70	0.44	0.64
		(2)	(3)	(5)	( $D = 0.067$ , $\chi^2 = 0.54$ , d.f. = 2)
2. Picture Preference (Diff.: $\chi^2$ )	Maori	0.90	0.80	0.60	0.77
	pakeha	0.85	0.75	0.45	0.67
		(—)	(—)	(—)	(1.82)
3. Doll Preference (Diff.: $\chi^2$ )	Maori	0.80	0.55	0.25	0.53
	pakeha	0.85	0.65	0.25	0.58
		(—)	(—)	(—)	(0.01)

It was concluded that a significant relationship between the *E* variable and performance on attitude tests was restricted to Maori Ss, and then only at the oldest CA level examined.

Finally, the present results point to the influence of the interviewer's ethnic grouping, in some circumstances, upon Ss' responses to tests of ethnic attitudes. In New Zealand, this effect was observed only among Ss who are members of an ethnic minority. On the last point further cross-cultural evidence is desirable. In any given instance, however, the provision of own-race *E*s should be considered an automatic control in ethnic attitude research.

This study was aided by funds made available by the Carnegie Social Science Committee of the University of New Zealand. The author is indebted to Mr Edward Bennett who acted as Maori experimenter.

#### REFERENCES

- DREGER, R. M. & MILLER, K. S. (1960). Comparative psychological studies of Negroes and whites in the United States. *Psychol. Bull.* **57**, 361-402.
- HEINE, R. W. (1950). The Negro patient in psychotherapy. *J. clin. Psychol.* **6**, 373-376.
- MORLAND, J. K. (1958). Racial recognition by nursery school children in Lynchburg, Virginia. *Soc. Forces*, **37**, 132-137.
- RANKIN, R. E. & CAMPBELL, D. T. (1955). Galvanic skin response to Negro and white experimenters. *J. abnorm. soc. Psychol.* **51**, 30-33.
- THOMPSON, R. H. T. (1959). European attitudes to Maoris: a projective approach. *J. Polynes. Soc.* **64**, 22-34.
- TRENT, R. D. (1954). The colour of the investigator as a variable in experimental research with Negro subjects. *J. soc. Psychol.* **40**, 281-287.
- VAUGHAN, G. M. (1962). The social distance attitudes of New Zealand students towards Maoris and fifteen other national groups. *J. soc. Psychol.* **57**, 85-92.
- VAUGHAN, G. M. (1963). Concept formation and the development of ethnic awareness. *J. genet. Psychol.* (in press).

- VAUGHAN, G. M. (1964a). Ethnic awareness in relation to minority group membership. *J. genet. Psychol.* (in press).
- VAUGHAN, G. M. (1964b). The development of ethnic attitudes in New Zealand school children. *Genet. Psychol. Monogr.* (in press).
- VAUGHAN, G. M. & QUARTERMAIN, D. (1961). Students' acceptance of Maoris: a structured picture test. *Percept. & Mot. Skills*, **13**, 190.
- VAUGHAN, G. M. & THOMPSON, R. H. T. (1961). New Zealand children's attitudes towards Maoris. *J. abnorm. soc. Psychol.* **62**, 701-704.

*Manuscript received 20 February 1963*



## Critical Notice

By A. R. JONCKHEERE

*Thought and Language.* By L. S. Vygotsky. London and New York: John Wiley & Sons, 1962. Pp. 168. 38s.

This is a version of the Russian edition which first appeared a few months after the author's premature death in 1934 at the age of 38. In preparing the book for publication Vygotsky tried to combine a number of separate essays, many of them previously published, into a coherent whole by dictating various passages during his last illness. Possibly as a result of the haste with which the book was prepared it tended to be both repetitive and diffuse, though the translators have tried to remove these flaws by abridging and editing the original text. Despite their efforts, however, the translation still shows traces of these faults, and the disjointed nature of the work together with Vygotsky's strong, digressive and polemical tendencies, make it extremely difficult to follow his at times brilliant and original trains of thought. As a consequence the content of the book is perhaps best presented by giving a brief summary of its separate chapters.

Vygotsky starts by considering the theoretical presuppositions which hindered psychologists investigating the connection between thought and language. According to him, they tended either to identify the two—and hence refused to face the problem of their interrelation—or to investigate them in isolation, trying to relate them later by using an 'external connection between distinct processes'. Vygotsky then introduces his fundamental methodological thesis, that progress can be made only if we analyse complex phenomena into 'units' and not 'elements'; for analysis into 'elements' leaves the investigator with the task of seeking 'the mechanical interaction of the elements in the hope of reconstructing, in a purely speculative way, the vanished properties of the whole'. 'Units', however, unlike 'elements', retain all the basic properties of the whole, and cannot be further subdivided without losing them. In the present context Vygotsky proposes the fruitful 'unit' to be 'word meaning', and claims that the very nature of the connection between word and meaning changes as a result of the child's general development and the way his thought functions: 'Thought and speech are not connected by a primary unchanging bond.'

In the second chapter, a version of Vygotsky's preface to the Russian edition of Piaget's first two books, Vygotsky criticizes Piaget's interpretation of the concept of egocentric speech. For Piaget, egocentric speech was a direct expression of the child's egocentric thought, and an intermediate stage between primary autism and the gradual socialization of thought as the child grew up. For Vygotsky, the sequence is virtually reversed. Egocentric speech is considered to be a phenomenon of the transition from the initial social activity of the child to his more individualized activity: speech for oneself, and later inner speech, originates through differentiation from speech for others. In a pamphlet entitled *Comments*, issued separately with this volume, Piaget agrees with most of Vygotsky's criticisms and indicates how much of his subsequent work, in the quarter of a century since the publication of his first books, is in substantial agreement with Vygotsky's viewpoint.

Next, Stern's theories concerning the child's sudden discovery that 'each thing has a name' are rejected for being too 'intellectualistic'. According to Vygotsky, Stern attributed this discovery to an almost innate tendency to search for meaning, and he dismisses the thesis in a typical passage: 'Stern answers the question of why and how speech acquires meaning by saying: from the intentional tendency, i.e. the tendency towards meaning. We are reminded of Molière's physician who explained the soporific effect of opium by its soporific qualities.'

In the next chapter Vygotsky analyses Köhler's and Yerkes' studies of chimpanzees, and infers that though these animals show the rudiments of both thought and language (mainly expressive), there is no evidence that at this level the two are in any way interrelated. This is chiefly because the conditions required for the ape's effective intellectual functioning are not those which would lead to the discovery of the functional use of signs or speech. He concludes that 'in the phylogeny of thought and speech, a pre-linguistic phase in the development



of thought, and a pre-intellectual phase in the development of speech, are clearly discernible'. Similarly, with very young children this relative independence of thought and speech can also be observed, following different lines of development until they meet at about the age of two, 'whereupon thought becomes verbal and speech rational'.

The fifth chapter analyses the process of concept formation in children. By using the blocks and the technique with which his name is uniquely associated in this country, Vygotsky distinguishes between three major developmental phases prior to the attainment of abstract conceptual thinking. At first the child can only form unorganized heaps based on vague subjective bonds having little relation to the objective characteristics of the blocks. Later these characteristics are taken into account, but only to produce 'complexes', in which the relations between the blocks selected are concrete and factual rather than abstract and logical. The bridge between 'complexes' and concepts is achieved when 'pseudo-concepts' appear. These are groupings having all the appearance of true conceptual classification, but which further examination shows to have been based on 'complex' thinking. Fascinating digressions into philology, the language of primitive people, and communication among deaf-mutes illustrate, and perhaps confirm, the distinctions Vygotsky tries to make between these phases of concept formation.

In the sixth chapter, the longest, most diverse and difficult to follow, a distinction is drawn between children's spontaneous and non-spontaneous concepts; an important example of the latter being the scientific ideas which are mainly acquired by formal instruction. Since spontaneous concepts develop as a result of everyday experience, it naturally follows that scientific concepts defined in terms of them must also develop. Vygotsky, however, further maintains that conversely, as scientific concepts develop, so do the spontaneous concepts upon which they are based, and concludes that: 'The development of the child's spontaneous concepts proceeds upward and the development of his scientific concepts downward, to a more elementary and concrete level.' An important, and perhaps surprising, pedagogical consequence is that, according to Vygotsky, instruction should not necessarily wait for the appearance of the appropriate spontaneous concepts. This general thesis is then used to analyse the complicated interactions between instruction in school subjects, and the child's thought processes peculiar to his developmental stage. There are, again, numerous digressions, one of the more interesting being the suggestion that a distinction be made between a child's mental age as assessed by the usual mental tests, and the level the child is capable of achieving in such tests with assistance. Vygotsky suggests that a score based on the discrepancy between these two mental ages will be a more helpful clue to the dynamics of intellectual progress than the customary mental age.

The last chapter has already appeared in English (*Psychiatry*, vol. 2, 1939) and, besides repeating many of the previous arguments and criticisms, contains further speculations concerning the origins and functions of inner speech. Observing that, provided the context is clear and there is sufficient rapport between speakers, public speech can be drastically abbreviated yet still perform its communicative function, Vygotsky suggests that inner speech is an extreme form of this tendency, leading even to the disappearance of words. For since one is then talking to oneself, rapport and context are at their most effective strength. This thesis is brilliantly illustrated by apt quotations from plays, novels, poems, and a discussion of Stanislavsky's techniques of theatrical production, while Vygotsky himself expresses it by remarking that, 'While in external speech thought is embodied in words, in inner speech words die as they bring forth thought.'

This bald summary hardly does justice to the wealth of perspicacious, subtle, and suggestive observations which are scattered throughout the work, indicating forcefully the extreme difficulty of the problems Vygotsky attempted to solve, and that present day psychology has made comparatively little progress since his time. In his attack on these problems Vygotsky discusses a vast range of phenomena, but an astounding omission is that he never considers aphasic, and gives only a passing mention of schizophrenic, language disorders. Neither Hughlings Jackson's nor Henry Head's works are mentioned, though they must surely have been known to him, and would have supplied valuable evidence for some of his ideas.

In addition to the polemical tone of the work, which tends to obscure Vygotsky's original thinking, the experimental studies are often so cursorily reported that at times it is almost impossible to gather what were the procedures he employed. When, for example, he gives an account in Chapter 5 of his famous work using the 'Vygotsky blocks', in order that the



reader might reasonably follow the discussion the translators were obliged to supply quotations (from Hanfmann and Kasanin's *Conceptual Thinking in Schizophrenia*) giving details of the experimental situation. It is unfortunate, moreover, that according to Professor Luria none of the experimental studies mentioned in the book were ever published in full. But the work must have been of seminal importance in the history of Russian psychology though, as Professor Bruner states in his Introduction, it was apparently suppressed in 1936 and only reappeared in 1956. It freed Russian psychologists from the domination of Pavlov whilst at the same time, by emphasizing the developmental aspects of behaviour, it enabled them to avoid some of the pitfalls of the Gestaltists.

Today we can appreciate how many of Vygotsky's ideas were later developed by philosophers and literary critics such as Wittgenstein, Ogden, Richards, Empson, and others concerned with the problems of language and meaning, but it is difficult to think of many psychologists who have since contributed as much as Vygotsky did within this domain: significantly, it is the first work by a psychologist to appear in the series 'Studies in Communication', the two previously published volumes being by Quine and Cherry.

## Book Reviews

*The Perception of Causality.* By A. Michotte. Translated by T. R. & E. Miles. London: Methuen, 1963. Pp. xxii + 425. 45s.

All psychologists must welcome the appearance, even if long delayed, of this very competent translation by Professor and Mrs Miles of Michotte's *La Perception de la Causalité*. Too little is known by English-speaking psychologists of the very extensive body of work carried out by Professor Michotte at the University of Louvain. Yet not only the topic studied, but also the general approach and methodology, constitute one of the most valuable and important contributions to psychology in the last twenty years. It is to be hoped that this English edition will be widely read and thoroughly pondered. It should be mentioned that it contains an entirely new section, not in the French editions, on recent experiments and certain new hypotheses based on their results.

But first, for those unacquainted with this work, a brief outline of how Michotte began his investigation of the manner in which the phenomena of mechanical causality may be demonstrated and experienced: In the initial experiment, a black square moves across a screen until its edge touches that of a motionless red square, and then stops; after a brief interval of time, the red square moves away in the same direction; observers then report that they perceived the moving black square to 'cause' the red square to move in turn—the so-called 'launching effect'. Clearly, no physical causality is involved; yet the perception of causality is spontaneous, provided that the temporal and spatial conditions are suitably arranged. It should be noted that Michotte carefully refrained from suggesting to his subjects that they might perceive these causal effects; and in all his later work he used naive subjects who had no expectations as to what they would see.

The conditions in which these and similar effects were perceived were investigated by Michotte and his collaborators in over a hundred experiments of a similar nature, using in the main the movements of simple shapes across a screen. In the course of these experiments it was possible to distinguish between effects such as the above, in which observers reported immediate and spontaneous impressions of causality; and those in which other types of inter-related movements were reported, including many which were similar to movements encountered in everyday life. Here the subjects reported that, for instance, it was 'as if' one movement triggered off another, or 'as if' a large object was carrying a smaller one upon it. In other words, a series of events for which there was no obvious explanation was described in terms of similar series of events experienced in everyday life. Michotte's contention was, however, that the initial causal impressions could not be explained in terms of familiar everyday life experiences. Indeed, causality was sometimes perceived in circumstances in which it would not ordinarily appear, as for instance when the black square did not actually touch the red square but was separated from it by a blank space. Thus although he accepts Piaget's observations that children gradually develop an understanding of the circumstances in which causality normally occurs through the movements they themselves 'cause', for instance by dropping or pushing objects, Michotte maintains that nevertheless we perceive mechanical causality spontaneously and without prior experience in certain well-defined physical conditions, just as we perceive movement.

This is, of course, a gross over-simplification of Michotte's theories and experimental results. It is impossible to pass judgment on these without careful reading of the whole book. Even in translation it is not always easy to grasp the subtlety of his concepts, particularly of the phenomenal nature of inter-related movement effects, nor to follow the arguments he has put forward to account for them. But although this work and the particular phenomena to which it is devoted may be said to be primarily of interest to students of perception, nevertheless the approach and methodology are of great importance in psychology generally. One must note firstly the extreme ingenuity of the experimental techniques, which employ very simple but carefully controlled presentations to demonstrate the conditions in which the different phenomena are experienced. Thus Michotte cannot be faulted by the criticism



often levelled at the Gestalt psychologists that their experiments were inadequately controlled; nor that the subjects knew beforehand what they were expected to see. Particularly significant is his use of the spontaneous verbal reports of naïve subjects as constituting evidence on the phenomenal nature of their perceptions. Clearly, evidence of this kind must be obtained and handled with care; but it has too often been neglected in behavioural studies. In the latter, the experimenter has assumed that he can infer the nature of the subject's experience from his reactions to the experimental conditions, with the consequence that conflicting results are often inexplicable. Moreover, in adopting the hypothetico-deductive method, the experimenter has often failed to allow for important types of experience not envisaged in his hypotheses, because the subjects have not been encouraged or even permitted to report them. There is a great number of phenomena which might profitably be studied by methods similar to these, observing the spontaneous percepts arising from simple and easily controlled perceptual configurations; for instance, simplified representations such as those of Brunswick of human emotional expressions. However, it is clear from Michotte's work that hypothesizing as to the nature and origin of such experience is likely to be a difficult and complex problem. There are few psychologists with the originality of mind, the ingenuity in experimental design and technique, the psychological insight and the expertise in argument, who could produce work such as this.

M. D. VERNON

*Research in Personality.* Martha T. Mednick & Sarnoff A. Mednick (Eds.). New York: Holt, Rinehart, and Winston, 1963. x+627 pp.

This volume brings together a cafeteria assortment of papers for the consumption of advanced undergraduate or graduate students in psychology. Over fifty journal articles (none published prior to 1950) have been assembled to give students a close view of the substance and method of contemporary research in personality.

In an introductory section the editors present a series of 'background papers' intended to acquaint the reader with aspects of current conceptualization in the field. (Among the more thought provoking of these are Cronbach's argument for a united discipline of psychology and Allport's brief for a broadened conception of psychodynamics.) In succeeding portions of the volume, papers are grouped to deal with the following: experimental studies of development, perception and personality, anxiety and stress, expression of conflict through fantasy, conflict resolution, personality organization, and illustrative examples of the cross-cultural and case history methods.

Although the editors themselves have written no interpretive comments to link separate selections, in their choice of items included they have generally coped with this issue in a satisfactory fashion. Review papers in a number of instances provide context and introduce specific areas of research. Further, critiques and replies to critiques supplement several of the research reports. Thus students are provided with front-row seats from which they can attend to successive statements in the research dialogue.

To an extent the volume suffers from the neglect of the broader social context as a significant factor in personality formation and functioning. With the exception of the two final papers in the collection, the relevance of socio-cultural factors has been largely ignored. Class and socio-economic status go unmentioned in the index. A further indication of this imbalance is the fact that of the 30 papers presenting original data on human adults, three-quarters of these use college undergraduate or graduate students as research subjects. The scientist-academician's tendency to tap that most readily available mine of natural research resources, his students, is understandable. And in numerous instances this predilection for using as subjects the more articulate and well-educated members of the public contributes no distortion to the basic phenomena being considered. Yet the cumulative effect of ignoring such personality-relevant factors of social differentiation as class and ethnic group membership limits the perspective of what is otherwise a useful volume on personality research.

JOHN D. CAMPBELL



*Psychology Through Experiment.* G. Humphrey (ed.). London: Methuen, 1963. Pp. 307. 30s.

It is difficult to consider seriously the possibility of even a general course in psychology being taught without students being introduced to practical and demonstration work of some kind. Even so authors have apparently been slow or unprepared to meet the needs of laboratory courses. The large research Handbooks cannot, of course, be expected to cater for less advanced or introductory experimental courses while the available smaller and cheaper texts usually either cover very limited fields only or are 'cook-books' which often insult the intelligence of our students. For these reasons one welcomes with appreciation the appearance of a book which has at once a strong practical orientation but which at the same time does not neglect the theoretical consideration of its concepts, hypotheses and interpretations.

Professor Humphrey, as editor, has contrived by his choice of contributors to indicate the width of modern psychology in a volume which, considering its size and cost, is probably more comprehensive than any other available. In his 'Introduction for Humanists' he attempts to clear up some of the misunderstandings which have arisen about such problems as experimenting on the human mind, the value of comparative studies, interrelationships between scientific systems, scientific method and the contributions of philosophy and psychoanalysis to psychology. It is unfortunate that he tries to do so much since his argument is inevitably cryptic and unconvincing. Significantly, neither Freud nor Wittgenstein get a fair hearing and it is perhaps worth noting that although the historian can only conjecture 'What song the Syrens sang' (p. 5) he is able to ascertain the direction of the winds which blew ships toward them!

Even the most casual glance at the rest of the book indicates clearly the skill and professional 'know-how' of each of the contributors. Mr Gregory's argument is often beautifully elegant and he combines the techniques devised for his own research with what probably, taken alongside the contribution of Professors Kay and Szafran, is the most easily understood description of 'information theory' available in the existing literature. Dr Dixon chooses three experiments, on apparent movement, constancies and word-recognition thresholds which illustrate some different approaches to perceptual theory extremely well. Professor Hunter again chooses three experiments on serial-remembering, retroactive interference (note not inhibition) and some effects of 'set' on remembering. These experiments, along with those described by Mr Foss, notable of which are the experiments on unconscious and motor factors in thinking, are probably more simple to conduct than any others in the book. Dr Watson selects four demonstrations on learning; conditioned hand-withdrawal, serial learning of nonsense syllables, animal maze learning and human problem solving, which indicate clearly that 'associative-connections'-type theories of learning require extensive modification. This contribution is particularly noteworthy since it includes an extremely well argued account of how to avoid artefacts in experiments of this kind. Dr Deutsch's contribution would be memorable if it were confined only to his advice on the care and use of laboratory rats. However, he also describes experiments on drive discrimination, spontaneous alternation and latent extinction in the course of which he indicates the development of his modifications of Hull's learning theory.

Readers, students particularly, may be forgiven for feeling that some contributors have not made enough concessions to them. Gregory, for example, uses the word 'perception' in a number of different ways (compare for example p. 33 with pp. 24, 34 and 45); Watson's argument and procedures for his conditioning experiment would probably have been simplified if the 'extinction' trials had been run first and a hypothesis formulated in terms of the temporal course of responses; there is a real possibility of misunderstanding in Hunter's logic relating to the Treatment of Results of his serial-remembering experiment (p. 149). The reviewer has not received lists of ancillary material available separately from the publishers and it is likely that students or tutors may wish to prepare their own lists. However, the argument concerned with the scoring of erroneously selected items only holds good if the recognition lists contain 30 items and when 10 items are presented in the learning lists. More generally, there is a surprisingly small amount of cross-reference and in view of the fact that the chapter on 'Statistics' was finally not included contributors should either have referred directly to P. G. Hoels 'Elementary Statistics' (see Introduction, p. 23) or, better, should have been afforded the very small amount of additional space necessary for illustrating



the wide reliance which they place upon statistical treatments, particularly non-parametric techniques. This latter procedure would probably have pointed up the need for the one really significant omission in the book—a chapter entitled 'Experimental Design' which one hopes Professor Humphrey will consider writing himself for future editions. However, important as these points appear to be in their detailed contexts, it should be emphasized that this is manifestly not a 'cook-book'. The onus of thinking is placed squarely on the reader who, after reading each chapter, must feel that he has been engaged in an important intellectual exercise—the questions left unanswered, the hypotheses left untested, the problems left unsolved remain for his further consideration.

GERALD H. FISHER

*Verbal Learning and Verbal Behaviour.* Charles N. Cofer (ed.). New York and London: McGraw-Hill, 1961. Pp. 241.

In recent years there has been a tremendous multiplication of research on verbal learning and verbal behaviour. This book, which consists of papers given at a conference in 1959 under the auspices of New York University and the U.S. Office of Naval Research, typifies many of the ideas and findings currently being discussed. There are, for example, papers on the present status of interference theory (Postman), individual differences in learning (Noble), the rôle of meaning (Bousfield) and the effects of frequency and transition probability (Deese). The speakers concentrate mainly on data and methods with little or no reference to any defined model of the learning process, other than concepts such as the S-R link, generalization, interference and mediation, which rather summarize than explain experimental findings.

The book contains many scattered points of interest: for example, Underwood's technique of modified free recall which clearly demonstrates and measures the unlearning of first list items during learning of a second list, and Deese's suggestion that word frequency may play no direct part in facilitating learning. Postman's reassessment of interference theory is useful and important. However, advance in this field has been slow, and one senses hard straining to adapt or even discard old ideas of rote learning rather than finding any achieved conceptual breakthrough. Goss, for instance, discusses the learning of conceptual schemes, a topic of interest, but which is analysed mainly in terms of stimulus and response variables previously applied to nonsense syllables, which bear little new fruit in this context. Bousfield still appears to equate the meaning of 'evil' with the 'representational response' of saying the word 'bad'. There is little or no mention of information theory, of Chomsky's ideas on grammar, of recent research on short term memory and the decay theory of forgetting, of language learning in children or of its breakdown in aphasia. At the same time, there is considerable emphasis in the discussion by some at least of the participants on points where earlier theories of verbal learning are proving inadequate in dealing with syntactical variables, one trial learning, linguistic context, problems of encoding and the choice of the relevant unit in learning. The form in which the conference is given has some advantages: the papers were written in advance and the discussion is introduced by a prepared paper and followed by only a summary of the major points made by other speakers, thus avoiding some of the bathos which may be produced when spontaneous comments are given in full. The discussion, indeed, often proves more stimulating than the paper itself. The book in general drives a few more nails into the coffin of a simple associationist analysis of verbal learning and behaviour, without giving birth to any major new understanding.

ANNE TREISMAN

*The Canford Families: A Study of Social Casework and Group Work.* P. Halmos (ed.). *The Sociological Review Monograph.* No. 6. January 1963. University of Keele. 25s.

According to temperament and mood, the social psychologist may feel overwhelmed or stimulated by the amount of work that urgently requires to be done. In all sciences there must be a continual stream of new questions to ask, answers to find. A special urgency is



felt, however, when the answers are eagerly awaited in the field of application. For the social psychologist the field of social work is one such.

In Canford a number of social workers set up a new form of co-operative framework in order to work with a set of sixteen families who were not 'problem families' as the term is normally understood, but yet had problems with which trained outsiders might help. The idea was broached in 1952, found financial backing in 1956-7, and was executed between 1957 and 1960. The monograph under review is a brief history of what happened, chapters being contributed by Elisabeth Howarth, Director of the project and tutor of the Child Care course at the North Western Polytechnic, Madge Hamilton and Moses Laufer, the group workers, Irene Spackman and Michael Power, the psychiatric social workers, and G. Stewart Prince, the psychiatric consultant.

Each chapter raises problems of theory and application for the social psychologist. The chapters on 'Aims and Development' and on 'Organization and Methods' are instances. What are the advantages and disadvantages of different administrative structures for co-operative social work? What are the adjustments which workers with diverse training and functions must, may, and will not, make in case conferences? Is a special set-up necessary if the functions described in this monograph are to be performed? If not, how many present facilities be guided to change in the desired direction?

The chapters on the families, their background and their problems use illuminating statistical tables and descriptions. Nevertheless, the workers, the commentators and the reviewer feel the lack of a theory of the family (family structures—family relationships—family processes—family in social context) which all would find equally useful for writing up their data and for discussing hunches or findings.

Again, although the accounts of the casework with the families are informative and rest securely on the experience of others elsewhere in the past, so that one feels that the workers know what they are doing, yet—for this reviewer at least—the feeling remains a feeling. There is no objective standard, either of logic ( $2 \times 2 = 4$ ) or of result (this-bridge-has-supported-that-load), to confirm the impression. This difficulty is especially relevant to the accounts of the group work. Boys were helped by one worker; girls and under-sevens by another. The two accounts differ in tone to give the impression that one was more successful than the other. Is this because one worker was better at the job than the other or because one was more sanguine in writing up the records?

How far we are from answering such questions! Yet if more money and more people were available, these problems are now within tackling distance. In conditions of greater affluence the decision to allocate funds for immediate help rather than for long-term thinking, evaluating and training would be less agonizing. Meanwhile, one is grateful to the Canford study, for building up the total stock of available facts, as well as on behalf of the area they served.

JOSEPHINE KLEIN

*Three Hundred Years of Psychiatry 1535-1860.* By Richard Hunter and Ida MacAlpine. Oxford University Press. Pp. xxvi, 1107. 84s.

*One Hundred Years of Psychiatry.* By Emil Kraepelin. London: Peter Owen. Pp. 163. 25s.

The long awaited publication of this anthology of psychiatric writings by Hunter and MacAlpine is a major event in the field of medical history. Until the last year only two major works existed in the English language—Hack Tuke's *A History of the Insane in the British Isles* (1882), and Zilboorg's *A History of Medical Psychology* (1941). Horrific details of the cruel treatment of lunatics, and an overdose of legislative history, make Hack Tuke difficult going. Zilboorg's partisan theoretical attitude vitiates what is otherwise a magnificent tour de force. Do Hunter and MacAlpine give us what we all long for—a scholarly, balanced, and wide ranging historical view of psychiatry? Unfortunately the method they have chosen at once makes for difficulty, and in some ways obscures their remarkable contributions to the subject. It is almost impossible to do anything but browse on this vast tome; consequently the growth



and development of ideas, and the personalities of the different authors of their books do not come out with the necessary clarity. If the method does not prove an insuperable difficulty to the reader, and relegate the book to the bedside—then what a wealth of information the reader will discover! Using over 300 selected English texts, the major part of the book is composed of extracts from these sources, with prefatory notes to each, either concerning the author or the subject. It is often alleged that 'British' psychiatry has never existed, that we were one of the backward races as far as psychiatry is concerned. How false this is can be judged from this survey of what is only a fraction of the literature. The sixteenth-century writers alone made a fundamental contribution to psychiatric thought when on the Continent of Europe the subject was still in thrall to demonological and theological ideas. A great deal of work still remains to be done in the field which is now being opened up by such magnificent publications as this. The distinguished mother and son who are its authors have produced a book which must be the envy of many of us who are working on the same subject.

In comparison, the reprint of Professor Kraepelin's book is a disappointing production, and one may ask why its translation was necessary? The epilogue by Dr Laqueur gives a highly erroneous account of some developments in modern psychiatry.

DENIS LEIGH

*The Unusual Child.* J. Roucek (ed.). London: Peter Owen, 1962. Pp. 293. 45s.

The central problem faced by the editor of a book, each chapter of which has been written by a different author, must be the question of how to instil into it a sense of coherence and consistency. There is little evidence that the editor of this book has faced this problem. He himself contributes a preface of only two pages where he indicates that the book aims at presenting 'a readable and all-inclusive summary of the available knowledge about a selected number of different categories of unusual children by outstanding authorities'. However, with a few exceptions (such as the chapters on 'The Emotionally Disturbed Child' by the Sherwoods, on the 'Socialization of the Atypical Child' by Elkin and on 'Community Services for the Unusual Child' by the Plattors) the writing is jargon-ridden and repetitive; and to claim that an 'all-inclusive summary' could be provided on any one handicap in a mere twenty pages (the average length of each chapter) is unrealistic as well as unrealized in this instance.

The following groups of 'unusual' children are dealt with: the intellectually gifted; the mentally retarded; the emotionally disturbed child; the speech handicapped; crippling conditions and special health problems; hearing problems; the blind, partially seeing and colour weak; the hard to reach child (dealing not with the autistic, as one might expect, but the pupil who refuses to learn); and children in need of institutional care; then follow some chapters on community services and vocational rehabilitation for the unusual child, and on parent and teacher education.

The editor implies that the book has been written for the 'average American teacher'. Not having any first-hand knowledge of the teaching profession in the United States, I cannot judge whether the average American teacher is really in need of such a general and oversimplified outline which also contains few practical suggestions as to actual teaching methods. Why the book should have been published in this country is even more difficult to see. The discussion is far too general for comparing teaching methods or administrative practices adopted in the United States and in this country respectively.

Superficiality and sketchiness also characterize the last chapter by Russell on the 'Provision for the Unusual Child in Western Europe'. Though it deals, in fact, with only two of the subgroups discussed in the rest of the book, namely with the subnormal and the gifted child, it does justice to neither. Out-of-date nomenclature is used (such as asylum or mental hospital when training centres and mental deficiency hospitals are referred to); the procedure for ascertaining subnormal children used in the County of Cheshire is described as being typical when in fact it is quite atypical as is the provision made in this county for the education of subnormal pupils; reference is made to Sir Cyril Burt, noting a connection between delinquency and mental deficiency, but later evidence to the contrary is omitted; as for the rest of Western Europe's procedures for ascertaining or educating the mentally subnormal child, the reader is left no wiser.



Lastly, there is only a combined subject and author index, which is both meagre and misleading, since only a few of the main topics and only some of the authors are cited. Each chapter is followed by a 'selected bibliography' and 'notes'; in fact, notes are provided in the first-named section whereas the second contains references only. Taken together, however, they provide a useful starting point for those who wish to pursue the study of any of the group of unusual children covered in this book.

M. L. KELLMER PRINGLE

*Signs, Signals and Symbols.* Stella E. Mason (ed.). London: Methuen, 1963. Pp. xii + 212. 35s.

The father figure of Dr Leopold Stein presides over this book which aims to present the British approach to Speech Therapy. The names of several of the best known speech therapists appear, Joan van Thal, Peggy Carter, Joan Pollit, Joyce Wilkins, Catherine Renfrew, among others. One lamented absence is Dr Muriel Morley, particularly as the stated aim of the book given force by Dr Stein is to follow the principles of science. In his paper he goes to great lengths to define science, and suggests renaming Speech Therapy 'Rhematology' as one of the steps towards establishing it as a science. But whatever weight Dr Stein may give to words and their usage, and no one would argue with the importance of this, what makes science is method, and most of the papers in this book can by no stretch be called scientific in this sense.

The book arose from a symposium on which the title is based. We are told that this is the first work to be published in England relating speech pathology to the wider study of human communication, and this is entirely to be welcomed. In the reviewer's opinion the title is singularly ill chosen. It has a grandiose flavour, and signals an expectation which is scarcely met by the content.

The quality of the papers is very uneven. The book contains contributions from people of professions other than speech therapy, including chapters by Professor Fry, W. Haas and J. L. M. Trim. There are papers on stammering, dyslalia, dysarthria, aphasia by speech therapists, many of which suffer by contrast with the former, with the notable exception of the work of the Nottingham group (Simms, Hartley, Grady) on dyslalia. G. Patrick Meredith's paper on 'The Quantum of Language' is the most interesting in the book, where a balance between mechanistic view of science and a humanist view is sought. It is indeed difficult in the fields of the natural sciences to strike the balance and most of the papers go nowhere near toward this.

In criticizing thus, the reviewer does not wish to suggest there is nothing of interest to be found in this book. It is not possible to do justice to several worthy papers that have not been mentioned, but if Speech Therapy or 'Rhematology' wishes scientific recognition, it must learn to apply the methods of science where possible.

BERENICE KRIKLER

*Gruppenstruktur und Gruppenleistung.* By H. Fischer. Bern and Stuttgart: Hans Huber, 1962. Pp. 136. DM23.80.

This book, *Group Structure and Group Performance*, reviews briefly for the German reader some of the types of mathematics which may be used to analyse the flow of information in experimentally constructed small groups, drawing on Lewin, Bavelas, Leavitt, some matrix algebra, some graph theory and some simple information theory. Those who read German more easily than French may find the pages introducing Berge's development of graph theory useful, and not all will be familiar with Flament's experiments on the effects of group structure. There is no index.

Mathematicians are urgently needed in small group studies and the neighbouring field of organization theory so that the structure of complex organizations may be more accurately described; possibly this appetizer may tempt some of them.

JOSEPHINE KLEIN



## Effects on Sociometric Status of Institutional Pressure to Adjust Among Retarded Children\*

By ROBERT A. DENTLER AND BERNARD MACKLER

*Teachers College, Columbia University*

This paper reports findings from a study of the social effects of pressures to adjust to a new group environment, as these were felt by newly arrived boys in a state hospital and training centre for mentally retarded children. Previous research reveals a correlation between cottage status and mental ability, and status and conformity to institutional norms. This study examined processes underlying these relations. For 29 retardates, mean age 9.6 years, mean I.Q. 56, group status initially correlated highly and positively with mental ability, social initiative as measured by observed frequency of attempted contacts with peers, and restriction of conduct as measured by frequency of aide discipline. In the second month of cottage life, following severe restriction by aides, these relationships changed: the abler, hence more frequently restricted boys, experienced a decline in status. Thus the effect of the institution is to control social relations by restricting activity and so modify group structure as to make relative ability, social or mental, a liability. Initially, the function of institutionalization is to conserve individual disabilities.

### INTRODUCTION

The research reported in this paper investigated the social effects of pressures to adjust to a new interpersonal environment, as these pressures were experienced by newly arrived children in a state hospital and training centre for mentally retarded children. Research reports by others have identified the correlates of status among retarded children (Marden & Farber, 1961; Dentler & Mackler, 1962a). If it is true, as these studies suggest, that the status structure of groups of institutionalized retarded children reflects the functions and norms of the institution, then one should be able to observe changes in the status relations of newly arrived children as they adapt to such norms.

The general problem is one of extending knowledge about the correlates of social status among children. Moreover, the study of social status among retarded children is of particular importance. It is a way of investigating features of group structure that do not depend for their emergence upon normal human capabilities for making fine symbolic distinctions; and it is a way of studying the larger problems of the impact of institutions on children that is partially free of the special complications introduced by mental disorders or criminal careers.

### HYPOTHESIS

The organizing hypothesis was that institutionalization would modify the social status and interpersonal relations of a newly arrived group of children in a state hospital and training school. Previous research (cited above) suggested that these

\* This study was conducted in part with funds from the U.S. Public Health Service Grant No. OM-111, and in part from the University of Kansas Research Fund.



children would differentiate their social positions somewhat in accordance with differences in mental abilities. This hypothesis was tested. Primary attention was directed, however, to the problem of the effects of institutionalization on their status relations.

The present analysis is thus directed at these questions: (1) Were positions in the sociometric structure associated with differences in mental ability? (2) Which boys were most frequently restricted by aides, and with what effect on their status in the peer group? (3) Did increasing restriction by aides, taken as an indicator of institutionalization, modify the sociometric structure?

### PROCEDURES

The 29 new boys, inhabitants of Cottage X, arrived at the institution within a few days of one another. After all were in place, the authors began 3 weeks of direct observation of interaction among the boys and the aides. Every hour of the day and night was covered, but emphasis was given to the waking hours when the boys were assembled as a unit. The authors stationed themselves in corners of rooms, tagged along to meals, to the playgrounds and gymnasium and to diagnostic sessions. As non-participant observers, they worked to gain a comprehensive impression of the pattern of cottage and institutional life. At the same time, they attended to the social behaviour of each boy, sampling sequences of individual, pair and subgroup activity as these unfolded in different sections of the cottage. Following every 2 hours of observation, the observer retired to tape record and transcribe a qualitative report of events.

In addition, the following measures were obtained for individual children:

1. *Ability.* To secure a global index of functioning abilities, four tests were administered to each of the 29 boys during the third and fourth weeks after arrival. These included the Porteus Maze Test of intelligence, the Parsons Language Sample (a test of verbal comprehension and vocal expressive abilities), a Cottage and Grounds Map Test, and the Parsons Index of Social Maturity (a laboratory test analogous to the Vineland Social Maturity Test which assesses independence training).

The nature and interrelation of these tests have been reported elsewhere (Dentler & Mackler, 1962b). They yielded intercorrelations in the range of 0.55 to 0.80. Scores on each instrument were normalized as T-scale equivalents and combined into a single index of ability. Mental ability as defined in this study is operationally specific despite the global character of the index. The index estimates level of abstract maze problem solving ability, language ability, cognitive awareness of the layout of cottage and institutional grounds, and basic social independence, characteristic of each child relative to the other children in Cottage X shortly after admission. The aim was to select measures of performance that had a bearing on adequacy of social behaviour.

2. *Initiative.* To obtain a comparatively direct measure of the interaction tendencies of each boy, an independent assistant trained in psychology coded the taped and transcribed observations in terms of the frequency of social acts initiated toward other persons by each child during the three weeks of cottage observation. Each social act or contact, friendly or hostile, was given an equal weight of one. The summed frequencies were normalized as T-scale equivalents.

3. *Restriction.* All children in this institution are controlled through a system of rewards and punishments. Clear, fairly universal regulations of norms have been established by the staff and aides. Aides file 'Pink Slip Reports' every time an incident occurs in which a child violates one or more of these regulations or resists correction sufficiently to elicit disciplinary action. The report details the nature of the violation, the disciplinary action taken, and the way the incident was terminated. These reports are filed in the child's records.

A frequency count was made of the number of Pink Slips each boy in Cottage X received each week during his first 21 weeks in the institution. As the correlations were significant from week to week between number of slips filed per child, for the present analysis the number of slips per boy for the first month, and for the number for the 4 weeks in the second month after arrival, were summed and normalized as T-scale equivalents.



Inspection of the Pink Slips as well as direct observation indicated that most disciplinary actions consisted of *restricting* the violator to a chair, to a corner of a room or to his bedroom in the cottage for a period of time that varied from 10 minutes to 2 hours. Extremely severe violations were disciplined by temporary placement of the child in the institution's 'disturbed ward'. Discipline also involved restriction of privileges such as use of play equipment, but the primary type of restriction was limitation of *social* activity.

4. *Status*. Major interest centred in the interpersonal relations among the 29 boys. Tests of sociometric status were devised which retarded children would understand. Each child's photograph was mounted with pins on a large portable board. The photographs were arranged randomly in six rows of five columns. At the close of their third week of observing, the authors established rapport with each child, one at a time, and administered the following questions in the privacy and comfort of a cottage bedroom:

Which boys would you most like to play with in the Day Hall?

Which boys would you most want to be (like)?

Which boys would you most like to work with in cleaning up the cottage?

The children responded with pleasure and apparent understanding to these questions. A minimum of *three* choices was elicited on each criterion. As a small internal check, the picture of a boy who did not live in Cottage X was included. Each child was asked to point to the picture of the boy who 'Does not live in your cottage'. Twenty-four of the children identified this picture correctly. Incidentally, children were invited to *point* to their choices, eliminating the problem of unscrambling a verbal response. Two weeks later, in the second month, these sociometric questions were administered again by a different tester. The arrangement of the pictures was changed and the tests were administered not in the cottage but in the hospital research unit.

From these separate administrations, two measures were devised. The number of choices received on each criterion was transformed into a normalized score. Status One and Status Two thus measure social acceptance as play and work associates and as objects of personal identification. Ranks on the three criteria of choice were significantly concordant for Status One ( $p < 0.05$ ) and for Status Two ( $p < 0.05$ ), using Kendall's W. The T-scale distributions on both status measures revealed excessive clustering about the mean. For this reason, the present report analyzes these data by classification of subjects into subgroups according to level of choice status.

## FINDINGS

The subjects, all boys, ranged in age from 6 to 12, with a mean of 9.6 years. Mental ages on the Porteus Maze ranged from zero (for three presumably untested subjects) to 15.5, with a mean of 5.5 years, and a mean I.Q. of 56. Staff professionals at the institution had tentatively identified these boys as educable. All but five of them (who were diagnosed by the institutional staff as exhibiting some degree of noteworthy central nervous system pathology) met Sarason's criteria for retardation as opposed to defectiveness (Sarason, 1959). Mental age scores on the Vineland Social Maturity Scale ranged from 2.5 to 9.6 years, with a mean of 5.5. Most of the boys had attempted and failed to cope with the academic and social demands made by public and elementary schools.

The 29 new arrivals were housed in a newly constructed, commodious dormitory cottage. This contained a large living room or lounge, a day hall with adjacent kitchen, two large bathrooms, and sixteen two-person and two four-person bedrooms, and smaller utility rooms. The cottage was located between two identical cottages in a semi-circle of dormitories surrounding the hospital, administration buildings, cafeterias and treatment and training centres. The cottage constituted a self-contained social system composed of the boys, eight to eleven aides assigned in pairs of 8-hour shifts around the clock, and infrequent visitors such as nurses, the



chaplain, and institutional staff members. The children moved periodically between the cottage and other structures in the institution, to eat, to recreate, to be diagnosed, treated and instructed. As new boys, however, they were not absorbed immediately into the life of the institution. Officially, they had been admitted for probationary observation, diagnosis and classification. Less Officially, but more significantly for the lives of the children, the first months of institutional life were set apart as a period of instruction in an adjustment to the institutional environment. The new boys were socialized toward institutionally acceptable performance of elementary yet (for retarded children) demanding routines of feeding, clothing, bathing, recreating and sleeping. Suffusing every repertoire of routine socialization was control through the imposition of restrictions.

As the primary agents of socialization, the aides of Cottage X spent much of their time regulating the behaviour of children. Aides were constrained to work toward rapid and effective socialization of the boys into the basic routines for their own work became less demanding or stressful to the extent that the new children learned to behave properly. The boys in Cottage X, moreover, had been classified as *educable* and therefore as capable, relative to other children, of effective learning. The aides were expected to succeed.

The frequency of efforts by aides to induce compliance among the new boys *increased* steadily during the first 8 weeks of cottage life. The evidence from 'Pink Slips' on pressure by restriction suggests that it reached a peak between the first and second months. In the second month it was reduced as the boys began increasingly to comply.

One effect of the attempt by aides to discipline the boys through restrictions toward compliance with institutional routines, was modification of the otherwise spontaneous character of interpersonal relations among the children. Observation suggested that continuous, increasing pressure through social restriction induced pressure toward withdrawal from social contact, avoidance of co-operation as well as conflict and competition among peers, and behaviour best described as socially cautious, submissive or, for some boys, deviously resistant to cottage codes.

In the first 2 months, boys in Cottage X were confined to the area of the cottage for roughly 15 out of every 24 hours. Few structured outlets for energies, particularly for social energy, were available, and these soon lost all novelty. Faced with the tasks of occupying time harmlessly and of training toward the elementary routines, the aides, with their limited resources, aimed at controls guaranteeing 'peace and quiet'.

The intercorrelation between ability, initiative, restriction in the first month and restriction in the second, are reported in Table 1. The four variables are significantly and positively associated. The correlation between ability and restriction increased substantially from the first to the second month, as did the correlation between initiative and restriction.

Consistent with this, the boys who were restricted most frequently the first month tended to be most restricted later. The correlation between first and second month restrictions ( $r = 0.56$ ) is limited by the fact that the more severely retarded boys resisted the demands of aides during the first few days after arrival but then



adapted to their new situation through withdrawal and the avoidance of conflict. The more able boys, in contrast, continued and in some cases increased their resistance to the demands of aides from arrival until the middle of the second month. In any event, the most able and the socially most active boys were most resistant to institutional controls and were restricted most frequently by cottage aides.

Table 1. *Intercorrelation of antecedent variables\**

	Initiative 2	Restriction 1 3	Restriction 2 4
1. Ability	0.47	0.35	0.49
2. Initiative		0.51	0.73
3. Restriction 1			0.56
4. Restriction 2			

\* Pearson  $r$  used, with scores normalized on T-scale.

$r \geq 0.35$ ,  $p < 0.05$ .

Table 2. *Correlations (contingency) of status one with ability, initiative, Restriction One and Status Two*

Boys ranked on	Association with Status One	
	C	$p <$
Ability	0.53	0.01
Initiative	0.56	0.01
Restriction One	0.38	0.10
Status Two	0.46	0.05

The boys were divided on sociometric status into three subgroups. Fifteen were classified as having average status scores in the first month. Eight were assigned to the low status group and six to the high. Association between levels of status and ability, initiative, restriction in the first month, and rank on sociometric status in the second month are reported in Table 2. Ability and initiative are highly and positively associated with status level in the first month, and restriction tends toward positive association with status.

Initially, then, the boys in Cottage X erected a status structure in which the least retarded, socially most active and least conforming boys were overchosen; and the most retarded, socially most withdrawn and most conforming were underchosen (Fig. 1).

Rank on status in the second month associated positively and significantly with status level in the first month (Table 2). Closer inspection of the raw data, however, revealed a pattern of changes in status levels deserving of further analysis. For example, among the six boys with highest status the first month, three dropped to average status the second month. Among the eight with lowest status initially, four changed from low to high status and one changed to average status. Five of the 15 boys with average initial status changed; but these five all declined to low status. Particularly crucial were the effects of regression toward average status among the overchosen boys, and movement to high status among half of the underchosen. The central portion of the status structure remained quite stable but important shifts occurred at the extremes.

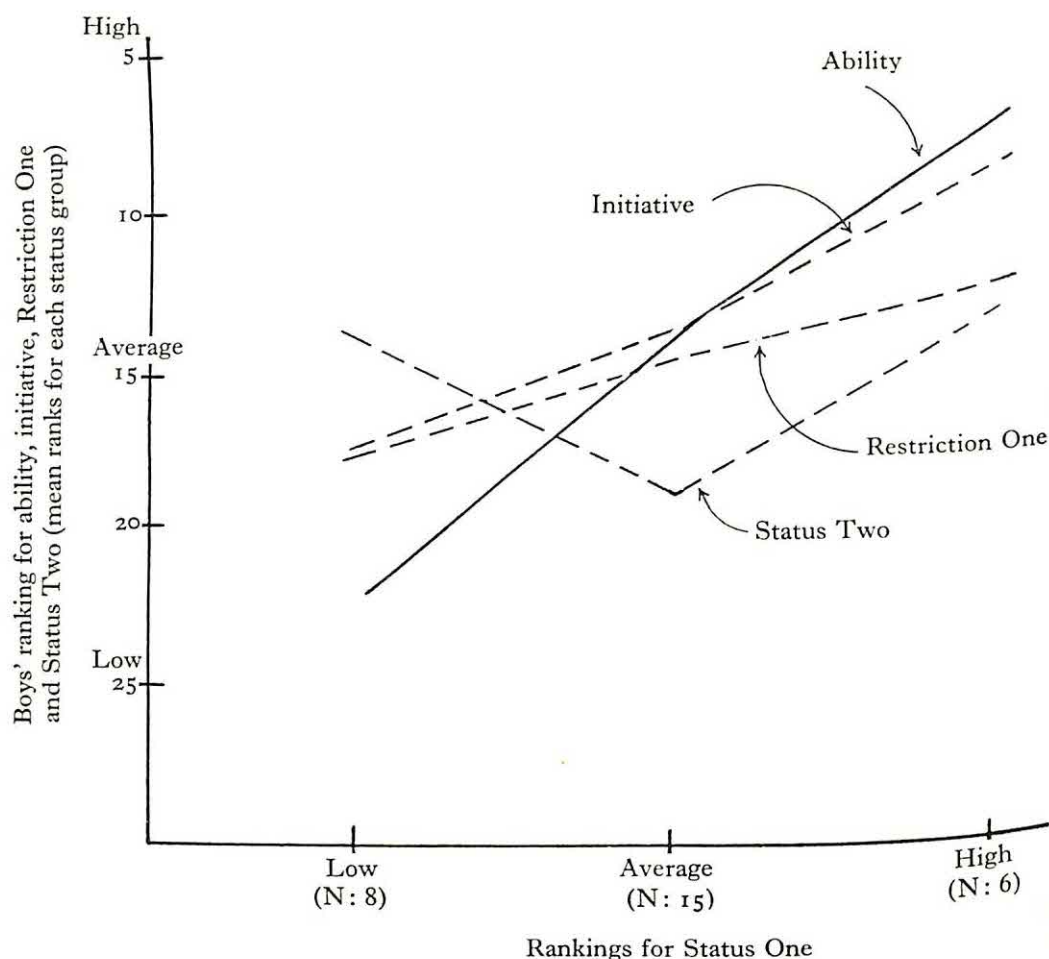


Fig. 1. Ranking of boys on Status One and ability, initiative, Restriction One and Status Two

Table 3. Association of status change with ability, initiative and change in restriction\*

Boys ranked on	Degree of association with status change		Nature of association
	C	$p <$	
Ability	0.73	0.001	Higher ability, greater status loss
Initiative	0.24	0.50	Unrelated
Restriction Change	0.61	0.001	Relative increase restriction, greater status loss

\* Contingency coefficient C indicates only degree, not nature of association. Nature taken from inspection of data.

In an effort to account for these changes, the boys were classified into three subgroups on the basis of magnitude of change in status. The 'No Change' group was kept large (N: 17) to control for changes that were little more than errors in measurement or artifacts of statistical regression. The seven boys whose scores reflected a gain in rank of more than five points were classified as high gainers. Five boys re-



flected a decline in status rank of ten points or more, and were classified as high losers.

Ability and changes in frequency of restriction were highly associated with changes in status (Fig. 2). The mean ranks of subgroups indicate that the most mentally able boys declined in status while the least able increased their status. When the boys were ranked from greatest to least increase in frequency of restriction

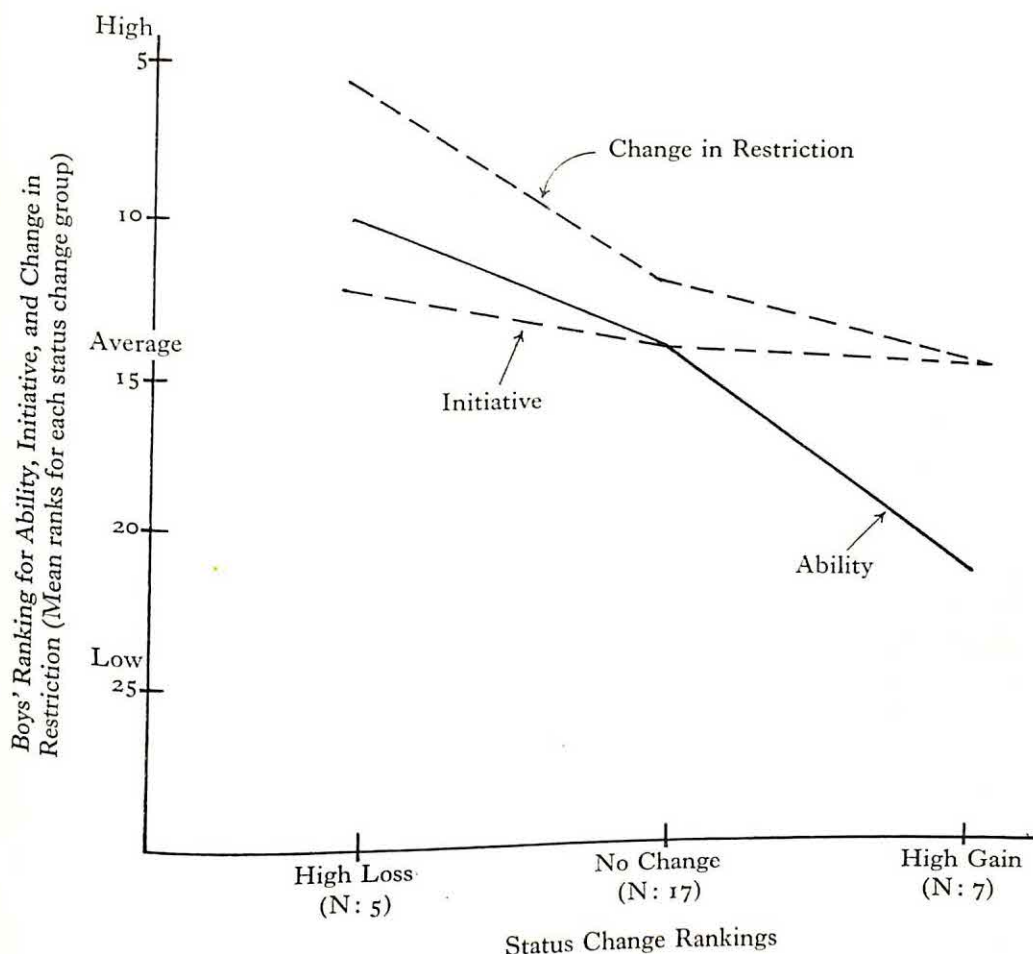


Fig. 2. Ranking of boys on status changes and ability, initiative and change in restriction

by aides, those who had dropped in status tended to be most frequently restricted. Initiative failed to discriminate between subgroups classified by changes in status. Another way to make this point is to note that the mean I.Q. of the high gainers was 0.47; the mean I.Q. of the high losers was 0.71.

#### DISCUSSION

The findings concerning status in the first month confirm previous studies of the relation between ability and status. These show that social positions within children's groups are distributed correlative with differences in ability to perform as a group member. Other factors affect the status structure, but generally the most able

individuals are also most capable of attracting the attention of their peers. More children are likely to identify with, or to attend to, peers who perform best as problem solvers or as sources of social and emotional gratification.

In view of the relation between social initiative and status, this study confirms and amplifies previous research. Where other measures of ability have been restricted principally to measures of intelligence, especially in research on retarded children, this study reveals that level of intelligence, vocal and verbal ability, social maturity, and frequency of interaction are all correlates of social status.

Cottage observations suggest that the boys who initiated more social interaction with their peers were socially more *visible* participants. In one sense, the measure of initiative, drawn as it was from a coding of the observations, is an artifact of this condition. The socially more active boys drew the attention of the observers, just as they made greater claims on the attention of their peers and just as they were most frequently disciplined by the aides.

The relationships established in the first month in Cottage X show that on arrival children tended to distribute their social sentiments (considered here as elements of reinforcement in a social structure) in a social psychologically 'rational' manner. Those peers who had least to contribute to group life and those who had most, were rewarded proportionately. This constitutes an extension into the institution of a model of status structure shared by most children in our culture.

Over the short term at least, one major *effect* of institutional procedures as reflected in disciplinary restriction was to penalize ability. Status rewards were re-directed, away from the most able boys and toward those with lower levels of intelligence, language skills and social maturity. Alternately, as some of the most able and active boys experienced increasing rejection by their peers, they increased their social activity to the point where they were disciplined increasingly by aides. In short, increasing resistance to cottage regulations could result from declining status.

The two variables, changing status and changes in restriction, probably interacted reciprocally. As the most able boys experienced changes in the social sentiments directed toward them—as they experienced a change from preference to avoidance—they probably intensified efforts at dominance and contact which resulted in increasing restrictions by aides and further avoidance by peers. As the most retarded boys learned to avoid discipline through withdrawal and passivity, they also experienced increasing preference from peers. Where the emphasis should be placed in the process cannot be answered from these data.

The change in social roles resulting from changes in occupancy of status positions is illustrated in the case of the group's most retarded member. This child remained the least disciplined or restricted boy in the cottage. His sociometric status, however, changed from twenty-third among 29 to fourth. The recorded observations reveal a change in his group role from an autistic, withdrawn and peripheral member to a person defined as the pivotal object of attention among both boys and aides. He was 'played with' and 'cared for' rather than becoming a play or work mate. In contrast, the most able member of the group changed from the fourteenth to the third most frequently restricted boy, at the same time he declined in status from the first



to thirteenth place. The recorded observations indicate that in this interval his role in the group changed from that of sought-after playmate to that of a cautiously avoided and socially isolated individual.

Without assigning antecedent influence to restriction, then, the interaction between adjustment to cottage life and social positions in the peer group was such that the status structure was modified. Boys were high and average initial status declined, and the boys with low status were elevated. Another aspect of this change has been reported elsewhere (Dentler & Mackler, 1961). Briefly, the authors found that the structure of reciprocal preferences was virtually eliminated between the first and second month. From a group with some 40 mutual ties, Cottage X boys moved toward a structure containing three *mutual* ties in the second month.

The authors interpret the overall process of the retarded child's initial adjustment to institutional life as follows: Social status is associated progressively with degree of individual compliance with institutional norms as these are applied by cottage aides. The character of this compliance is such that socio-emotional restriction or withdrawal is induced, contributing to a slow, relative decline in level of intelligence, language skills and social maturity. Failure or refusal to comply leads to reassignment to a more closely supervised cottage or to release from the institution. Reassignment is more likely, given the family situations from which such children originate. Closer supervision involves increased restriction and a subsequent further decline in functioning ability. In either event, initial mental and social abilities constitute relative liabilities for the individual child. Finally, in spite of this process, a low yet significant association persists between social status and mental ability among retardates as among normal children. For in the long term, the ablest retarded children learn to comply with, or deviate only covertly from, the most salient institutional norms.

#### REFERENCES

- DENTLER, R. A. & MACKLER, B. (1961). The socialization of institutional retarded children. *J. Hlth Hum. Behav.* **2**, 243-252.
- DENTLER, R. A. & MACKLER, B. (1962a). Mental ability and sociometric status among retarded children. *Psychol. Bull.* **59**, 273-283.
- DENTLER, R. A. & MACKLER, B. (1962b). The Porteus Maze Test as a predictor of functioning abilities of retarded children. *J. cons. Psychol.* **26**, 50-55.
- MARDEN, P. W. & FARBER, B. (1961). High-brow versus low-grade status among institutionalized mentally retarded boys. *Social Problems*, **8**, 300-312.
- SARASON, S. B. (1959). *Psychological Problems in Mental Deficiency*. New York: Harper Bros.

*Original manuscript received 15 January 1962*  
*Revised manuscript received 15 July 1963*

## **A Note on the Interpretation of the Guess Who Test in the Study of Sociometric Choice Behaviour**

By DERRICK PRITCHATT

*Department of Psychology, University of Leeds*

A previous finding of a high correlation between Guess Who and Sociometric Tests was confirmed, but this correlation obscures the fact that individual choices on these tests were not consistent. This was partly due to the fact that many of the characteristics of the Guess Who seemed to have little real significance to the children. However, even when they were allowed to give their own characteristics of a good friend, they still rarely attributed them to their own sociometric choices. It is argued that Guess Who characteristics can tell us very little about the reasons for sociometric choice.

### THE PROBLEM

For several years now the Guess Who Test initiated by Hartshorne, May & Maller (1929) has been used in conjunction with a Sociometric Test to investigate group relationships. The two tests correlate well, so that the Guess Who Test can be used to predict sociometric status (Eng, 1954); it has even been suggested that, for certain purposes, the Guess Who Test might be more useful than the Sociometric Test since 'it is slightly more indirect and the intent of the investigator is less apparent to the subjects' (Lindzey & Borgatta, 1954, p. 410).

The motives underlying friendship choices have also been investigated. Earlier workers (e.g. Potashin, 1947) have looked for correlations between sociometric choices and various characteristics of those people chosen. Thus, Thorpe (1955) found little or no correlation between sociometric choice and intelligence, neuroticism or age; on the other hand, Shukla (1951) found some connection between choice and common interests and 'companionship'—he gives a fairly long list of qualities that were considered desirable in a friend, including: humour, generosity, neatness, sportsmanship, kindness to animals and many others. From this type of work it has been an easy step to the use of the characteristics of the Guess Who Test to elucidate choice behaviour.

It has been maintained that the results obtained from the Guess Who Test will, when presented back to the children in conjunction with a Sociometric Test, give them 'a clear picture of which behaviour characteristics contribute to peer acceptance in their group' (Gronlund, 1959, p. 255). Bjerstedt (1956) also used the Guess Who Test to examine choice motivation. This does, of course, assume that the characteristics on this test have some real psychological meaning to the children in terms of friendship. The approach of the earlier workers who allowed the children to give their own characteristics rather than choosing them from a ready-made list would seem to avoid this assumption. The present study was an attempt to compare these two approaches with the actual choices made by individual choosers.



# METHOD

Both Sociometric and Guess Who Tests were given to two groups of co-educational grammar school children.\* The ages of the children were 12+ and 15+, and the sizes of the classes were 33 and 30 respectively. The sociometric test consisted of three choices on three criteria which were:

- (1) Seating preference for school work.
- (2) Preference for someone to take home to a party.
- (3) Preference for a companion to take to a film show.

Rejections were not asked for as experience has shown that there is often resentment towards such questions—this contention was substantiated by the reluctance expressed later by some children with regard to putting forward names for the less pleasant characteristics of the Guess Who Test. The Sociometric Test Sheet contained other questions, some of which were provided as 'padding' and some to throw light on issues beyond the scope of the present article. Among these questions was one requiring the child to list the three qualities that a friend should possess. The Sociometric Test was completed by the children before they received the Guess Who Test.

The Guess Who Test consisted of 20 enhancing and 20 detracting characteristics which were complementary and have been used by earlier workers, for example:

- 'Here is someone who is cheerful, jolly and very gay.'
- 'Here is someone who is often sad, worried and rather gloomy.'
- 'Here is someone who has no sense of humour and cannot see the funny side of things.'
- 'Here is someone who has a sense of humour and likes a good joke.'

The 40 groups of characteristics were presented in a jumbled order on the test sheet, and by the side of each one the child entered the name of one classmate that best fitted the description.

Answers to both tests were written, and the children were assured that their replies would be converted into code symbols by a person who would not even see them (my wife) so that their names would not appear on the data by the time it was examined. They appeared to be satisfied with this arrangement.

# RESULTS AND DISCUSSION

A rank order for the whole class was worked out for the number of choices received on the Sociometric Test. In the case of the Guess Who Test, choices received on each of the 20 enhancing characteristics scored as +1, and each of the detracting characteristics scored as -1: the algebraic sum was calculated for each child, and again a rank order was worked out for the whole class. A Spearman Rank Correlation Coefficient was now calculated between the rank order obtained from the Sociometric Test and that of the Guess Who Test. A coefficient of 0.74 was obtained for the 12+ age group and of 0.77 for the 15+ group ( $p < 0.01$  in each case). This result is in agreement with earlier studies.

The choices were now subjected to a detailed examination, however, and it was found that children who attributed positive Guess Who characteristics to a particular person did not necessarily choose him on the Sociometric Test. In fact, very few of these characteristics were attributed by a given person to anyone that they had chosen on any of the three criteria. Thus, the average number of characteristics, out of a possible 20, actually attributed to any first choice on any criterion was only 5 for the 12+ age group and 4 for the 15+ age group. Clearly, then, most of the characteristics held practically no significance to the chooser as far as his own choice was concerned.

\* I am indebted to the headmaster and pupils of Oldbury Grammar School for their kind co-operation.

An examination of the characteristics of a friend given by the children, themselves, showed them to be very varied, but 'loyalty-trustworthiness' accounted for the largest number of mentions (40 per cent of the total number, the next largest was only 18 per cent). Again, a check was made to see whether they had actually chosen the people, on the Sociometric Test, to whom they had attributed these characteristics. The only way to do this was to find out if they chose the person that they had picked for just those Guess Who characteristics that corresponded to their idea of the qualities of a good friend; fortunately, most of these qualities did appear on the Guess Who, particularly those most frequently given—these had a total of 142 mentions, as opposed to 42 mentions of characteristics not on the Guess Who Test.

The total number of first choices on all three sociometric criteria that had to be explained were: 53 for the 12+ age group and 57 for the 15+ age group.\* In the case of the 12+ age group, it was found that only 20 of their sociometric choices

Table 1

	Number of choices given to same sex (all criteria)	Number of choices given to different sex (all criteria)	Number of choices given to same sex (third criterion)	Number of choices given to different sex (third criterion)
12+ group	46	7	26	7
15+ group	45	12	18	11

were picked on the Guess Who Test as highest on one of the set of qualities that each individual considered to be desirable in a friend. We could at best, then, explain only 37 per cent of their choices. For the 15+ age group the corresponding number was 11, i.e. 19 per cent of their choices.

Choice on the third criterion (someone to accompany to the cinema) would seem to lend itself to a fairly straightforward interpretation of physical attraction—this interpretation, however, would be inadequate as seen from Table 1 which shows that most choices are given to members of the same sex. The near identity of columns 2 and 4 certainly suggests sexual attraction in these cases, but they are a small proportion.

### CONCLUSION

The important fact that emerges is that we can say very little about the reasons for choices. Though individuals put down characteristics of a friend, they often do not attribute them to the person they choose. It seems, then, that these characteristics do not greatly influence the child's choice and are simply socially acceptable values that they duly repeat. It is probable that only an individual approach will eventually reveal the particular needs operative in determining choice for a specific individual in a specific situation—the Guess Who Test is not a suitable tool for this purpose.

\* These figures arise as follows: if a child chose the same individual for all criteria, then there is only one choice to explain; if three individuals were chosen, there are three choices to explain etc.



REFERENCES

- BJERSTEDT, A. (1956). *Interpretations of Sociometric Choice Status*. Copenhagen: Lund.
- ENG, E. W. (1954). An Approach to the prediction of sociometric choice. *Sociometry*, **17**, 329-339.
- GRONLUND, N. E. (1959). *Sociometry in the Classroom*. New York: Harper.
- HARTSHORNE, H., MAY, M. & MALLER, J. (1929). *Studies in the Nature of Character. II. Studies in service and self-control*. New York: MacMillan.
- LINDZEY, G. & BORGOTTA, E. F. (1954). Sociometric measurement. In *Handbook of Social Psychology* (ed. Lindzey). Addison-Wesley.
- POTASHIN, R. (1947). A sociometric study of children's friendships. In *Personality and Sociometric Status* (eds. Northway, Frankel, Potashin). Beacon House.
- SHUKLA, J. K. (1951). A study of friendship. In *Studies in the Social Psychology of Adolescence* (ed. Fleming). Routledge & Kegan Paul.
- THORPE, J. G. (1955). A study of some factors in friendship formation. *Sociometry*, **18**, 207-214.

*Original manuscript received 1 May 1963*

*Revised manuscript received 30 June 1963*

## A Control for Social Desirability in a Semantic Differential

BY MARGERY H. KRIEGER

*Veterans Administration Hospital, Seattle, Washington*

A control for social desirability was introduced into a semantic differential appropriate for describing person concepts. The modified instrument consisted of 18 scales balanced for SD and four unbalanced evaluative scales as marker items. Two subject groups, 52 psychiatric patients and 40 non-patients, described *a doctor*, *a patient* and *myself*. Data from each group yielded six orthogonal factors, with evaluative factors loaded primarily on the four unbalanced scales. Comparisons of factor structure by group and by concept showed much less factorial correspondence between these English-speaking groups, and a higher degree of subject-concept-scale interaction, than is usually found with the semantic differential. Control for SD seems to make the scales more sensitive to subject and situational differences.

Recent research on personality inventories, attitude scales and similar paper-and-pencil measures has demonstrated that a subject's responses and resulting scores can be determined by many factors other than the content of the item and the degree to which the subject 'possesses' the trait presumably measured by it. A number of such factors, usually called response sets, have been identified and investigated since two early papers by Cronbach (1946, 1950) pointed up their potential importance. For instance, the effect of an acquiescence response set, or the tendency to agree (or disagree) with a statement regardless of content, has been demonstrated in several instruments (Bass, 1955; Jackson & Messick, 1957; Couch & Keniston, 1960). Similarly, a social desirability (SD) response set, or the tendency to agree with or endorse socially desirable statements and to disagree with or fail to endorse socially undesirable statements has been investigated extensively and found to operate in many personality inventories, Q-sorts, self-concept measures, etc. (Edwards, 1957; Cowen & Tongas, 1959).

It is possible to regard such response sets either as constant errors or as meaningful (and measurable) personality characteristics. Some instruments have been found to be so heavily saturated with such 'non-test-relevant response variance' (Marlowe & Crowne, 1961) as to be meaningless as measures of anything else (e.g., Cowen & Tongas, 1959). It is now clear that an inventory or scale intended to assess personality, attitudes, values, self- or interpersonal perceptions, etc., must in some way control or evaluate the effect of such sources of variance.

The present study introduces a control for SD response set in a semantic differential that can be used for the description of people. The semantic differential is a measurement technique introduced by Osgood and his associates (Osgood, 1952) in which selected concepts (e.g., self, other persons, objects, ethnic groups, abstract concepts, etc.) are rated on a series of seven-step scales defined at their ends by adjectives of opposite meaning. Typically, the subject is presented with a concept, in the form of a noun, and a number of descriptive scales which he checks to indicate what the concept connotes to him, much as follows:



## A PSYCHOLOGIST

strong	.....	weak
happy	.....	sad
good	.....	bad

The technique has been used in a variety of research situations, including studies of attitude (Nunnally & Kittross, 1958), personality change (Osgood & Luria, 1954), identification (Lazowick, 1955), and interpersonal perception (Helper, 1958); that is, for exactly the sort of measurement in which response sets such as SD should be evaluated or controlled.

Semantic differential concepts can be compared when they share a common semantic space defined by a common (or similar) set of underlying factors. Most factor analyses yield two or three major factors, no matter what subjects, concepts or scales are used. These factors can be usually be termed Evaluation (good-bad), Activity (active-passive) and Potency (strong-weak) or a combination of the latter two in a factor termed Dynamism (Osgood *et al.*, 1957). The evaluative factor is usually the largest, accounting for as much as 50 per cent of the total variance of the instrument and as much as 70 per cent of its reliable, or common, variance. A high degree of factor similarity, or semantic generality (Osgood, 1962), across concepts is the usual finding, even when comparisons are made between subject samples drawn from culturally and linguistically different populations (Kumata & Schramm, 1956; Triandis & Osgood, 1958; Suci, 1960). The factor structures of single concepts are more difficult to equate, and there is usually a good deal of concept-scale interaction (Osgood *et al.*, 1957). The possibility of subject-concept-scale interaction has not been investigated, and requires the separate factor analysis of single concepts for different subjects or subject groups.

The usual procedure for constructing a semantic differential is to select scales which have been shown to be more or less pure representatives of each of the major factors, including in the instrument an equal number of scales for each factor. With this instrument, subjects then are instructed to rate concepts determined by the particular hypotheses under investigation. Frequently some measure of association such as product moment correlation or the generalized distance function,  $D$  (Osgood & Suci, 1953), is calculated to reflect assumed similarity or dissimilarity between concepts. Bronfenbrenner (1958) has pointed out that estimates of assumed similarity in such designs can be distorted by the operation of response sets such as favourability. Semantic differential scales are seldom pure representatives of particular factors, and most of the scales in common use have rather clearly favourable and unfavourable ends. A scale such as strong-weak, often used to represent potency, can have a sizeable amount of its variance determined by evaluation, especially when the concepts being rated are people. With an evaluative factor accounting for most of the reliable variance in a semantic differential, a subject's set to describe concepts in socially desirable (or undesirable) ways could largely determine the size of indices of assumed similarity between these concepts. The present study has two purposes. First, it is an attempt to introduce a control for SD response set in a semantic differential designed for the description of person

concepts. Secondly, in order to determine the effect of this control on the structure of the instrument, it reports the results of factor analyses performed separately on the data from two subject samples describing three concepts. This permits comparison of factor structure between subject groups and between concepts on the same instrument, a procedure which hitherto has not been reported for a semantic differential. Findings as to factorial correspondence are compared with those of other studies.

### METHOD

#### *Scale construction*

From a pool of 1,200 adjectives appropriate for describing people, a sample of 200 was drawn randomly. Eliminating duplication and unusual or difficult words, a final sample of 100 adjectives was developed. These adjectives were rated by 145 undergraduate college students on a seven-point scale of social desirability from 'Very Desirable' to 'Very Undesirable' using Edwards' (1957) method. Median and Q were computed for each adjective, and five groups of adjectives of equal median favourability and minimal Q were formed. These were submitted to eight judges with instructions to extract from each group of adjectives as many pairs of antonyms as possible. Adjective pairings made by 50 per cent or more of the judges were selected for inclusion in the final instrument, a total of 18 in all. Each scale is thus balanced for SD since the bipolar adjectives defining the scale ends are matched for median favourability. They range from pairs of high to low favourability.

These 18 scales, together with four of Osgood's evaluative scales, included as marker items, constitute the semantic differential used in this study. The 22 scales are listed in Table 1. The four evaluative scales (good-bad, kind-cruel, foolish-wise and successful-unsuccessful) are, of course, heavily biased at one end for favourability.

Table 1. *Rotated factor matrix: Patient group*

Scales	Factors*						$h^2$
	Eval.	Dom.	Exp.	Lib.	Tens.	(Unnamed)	
Talkative-quiet	0.09	0.08	0.87	-0.05	0.05	0.02	0.78
Liberal-conservative	0.14	-0.34	0.21	0.66	0.25	0.09	0.69
Good-bad†	0.81	-0.10	-0.10	0.01	0.30	-0.26	0.83
Dominating-submissive	0.05	0.44	0.18	0.06	0.37	0.40	0.53
Practical-creative	0.53	-0.43	-0.05	0.04	0.02	-0.16	0.50
Passionate-controlled	0.08	0.10	0.59	0.48	-0.03	0.05	0.60
Strict-lenient	-0.09	0.34	-0.02	0.01	-0.10	0.68	0.60
Fragile-rough	0.62	0.11	0.08	0.04	-0.03	0.07	0.41
Kind-cruel†	0.76	-0.05	-0.36	0.29	0.09	-0.03	0.80
Cautious-confiding	0.37	0.10	-0.57	0.02	-0.04	0.08	0.48
Masculine-feminine	0.69	-0.10	-0.32	-0.10	0.16	0.07	0.63
Shameless-prudish	-0.22	-0.75	0.06	0.04	-0.01	-0.04	0.62
Easygoing-aggressive	0.34	-0.10	-0.10	0.50	-0.04	-0.10	0.40
Foolish-wise†	-0.61	-0.09	-0.05	-0.31	-0.20	0.57	0.84
Simple-complex	0.38	-0.66	0.38	0.31	0.05	0.03	0.82
Compliant-assertive	0.03	-0.65	-0.10	0.21	0.05	0.13	0.50
Disciplined-free	0.28	-0.17	-0.24	-0.66	0.32	0.27	0.76
Successful-unsuccessful†	0.73	0.04	0.05	-0.23	0.00	-0.06	0.59
Tender-firm	-0.21	-0.24	-0.36	0.05	-0.10	0.72	0.76
Blustering-weak	0.45	-0.01	0.03	0.08	0.58	-0.21	0.59
Conventional-unusual	0.64	0.10	0.12	-0.39	-0.22	-0.10	0.64
Excitable-placid	0.41	0.12	0.19	0.02	0.33	0.64	0.74
Per cent total variance	21	10	10	9	5	10	64

\* Eval. = Evaluation; Dom. = Dominance; Exp. = Expression; Lib. = Liberalism;

Tens. = Tension.

† Marker items.



Table 2. *Rotated factor matrix: Non-patient group*

Scales	Factors*						
	Eval.	Dom.	Lib.	Tens.	Exp.	(Unnamed)	$h^2$
Talkative-quiet	0.33	-0.64	-0.06	0.16	0.30	-0.28	0.72
Liberal-conservative	0.03	0.22	0.71	0.06	0.10	-0.17	0.60
Good-bad†	0.72	0.01	0.01	-0.05	-0.44	-0.03	0.72
Dominating-submissive	0.01	0.83	0.07	0.12	-0.03	0.10	0.72
Practical-creative	0.02	-0.40	-0.64	0.36	-0.10	-0.06	0.71
Passionate-controlled	0.29	0.36	0.17	0.08	0.51	-0.08	0.52
Strict-lenient	0.09	-0.02	-0.59	0.01	0.44	-0.07	0.56
Fragile-rough	0.59	-0.08	0.15	0.01	-0.08	0.30	0.47
Kind-cruel†	0.81	-0.05	0.01	0.09	0.22	0.08	0.72
Cautious-confiding	0.10	-0.21	0.05	0.00	-0.31	-0.70	0.64
Masculine-feminine	0.05	0.15	-0.59	0.01	-0.65	0.06	0.80
Shameless-prudish	0.30	0.61	0.18	0.15	0.40	-0.06	0.68
Easygoing-aggressive	0.62	-0.45	0.32	-0.15	-0.03	0.03	0.71
Foolish-wise†	-0.78	-0.09	0.13	0.34	0.04	-0.11	0.76
Simple-complex	-0.38	-0.42	0.14	-0.34	-0.14	-0.01	0.48
Compliant-assertive	0.27	-0.72	-0.05	0.14	-0.08	-0.28	0.70
Disciplined-free	0.29	0.02	-0.33	0.82	0.04	-0.07	0.87
Successful-unsuccessful†	0.74	0.00	-0.17	0.19	0.10	0.04	0.62
Tender-firm	-0.06	-0.50	0.52	0.05	-0.07	0.39	0.68
Blustering-weak	0.01	0.15	0.54	0.36	0.13	0.09	0.47
Conventional-unusual	0.07	0.75	-0.43	0.00	0.17	0.09	0.79
Excitable-placid	0.21	0.16	-0.04	0.60	-0.05	0.49	0.67
Per cent total variance	17	17	12	8	7	6	66

\* Eval. = Evaluation; Dom. = Dominance; Lib. = Liberalism; Tens. = Tension; Exp. = Expression.  
 † Marker Items

### Subjects

Two groups of subjects (Ss) were used, an all male psychiatric patient group consisting of 52 in-patients from VA Hospitals, Seattle and American Lake, and a non-patient group of 40 Ss consisting of 32 members of a religious organization in Seattle and eight other adult professional persons. All Ss were English speaking, but the two groups differed in many respects besides the patient-non-patient difference. Forty-three per cent of the non-patient group were women. The mean age of the non-patients was slightly younger, and they were of higher educational achievement and socio-economic status than the patients.

### Procedure

Each S used the semantic differential to describe three concepts, *a doctor*, *a patient in a hospital* (with general hospital specified for the patient Ss), and *myself*. Scale order was standard, as was the order in which the concepts were presented.

### Scoring and treatment of data

Scale scores from one to seven were assigned. First, in order to compare the two groups, patient and non-patient, each Ss scale scores were totalled across the three concepts. Scores were combined in this manner so that the method of analysis would be comparable to that employed in other studies comparing groups (Kumata & Schramm, 1956; Osgood *et al.*, 1957, p. 224). For each subject group, product moment intercorrelations between the 22 scales were obtained using the combined scale scores. The two intercorrelation matrices were then factor analysed separately by the principal axes method, with the aid of the IBM 650. Each factor analysis yielded six principal axes factors which accounted for 65 per cent of the variance of the instrument. Graphic, orthogonal rotations to simple structure were then performed. The two rotated factor matrices are reported in Tables 1 and 2.

Second, in order to compare concepts, each of the three concepts was factor analysed separately for each group. The six 22-scale product moment intercorrelation matrices were

factor analysed by the principal axes method. In each analysis the first four principal axes factors accounted for about 50 per cent of the variance. In each case, after extraction of the fourth factors, residuals were generally low and indicated that any remaining factors were specific to single scales. Graphic orthogonal rotations were to simple structure.

#### *Predicted results*

An evaluative factor is to be expected in this instrument because of the inclusion of the four unbalanced marker scales. Complete control of SD in the balanced scales should result in a factor with high loadings restricted to these four scales.

Factorial correspondence can be estimated by indices of factor similarity between the factors of any two rotated factor matrices. The index of factor similarity ( $e$ ) developed by Wrigley & Neuhaus and discussed by Osgood *et al.* (1957), is a coefficient of proportionality, a correlation coefficient that does not include the point of origin. Like correlation coefficients,  $e$  can vary from 1.00 to -1.00. A resulting diagonal matrix of indices will reflect factorial correspondence to the extent that indices at the diagonals or between similar factors approach unity while all other values approach zero, that is, to the extent to which it approximates an identity matrix.

If the present instrument, with SD controlled, has the same semantic generality as the unmodified instrument, the first type of factor analysis described above (combined concepts) should yield factors which are highly similar when the two groups are compared. For the second type of factor analysis (single concepts), comparisons between concepts for each group are not expected to have a high degree of factorial correspondence, but descriptions of the *same* concept by the two groups should yield highly similar factors.

## RESULTS

### *Evaluative factors and the control of SD*

Reference to Tables 1 and 2, the rotated factor matrices for the patient and non-patient groups, shows that each analysis yields a factor which is clearly evaluative, identifiable by high loadings on the four marker scales. For the patient group (Table 1), the evaluative factor accounts for 21 per cent of the total variance of the instrument, and for the non-patient group, 17 per cent of the variance. In the patient group, nine, and in the non-patient group five of the remaining 18 scales have evaluative connotations (loadings above 0.30), despite the effort to balance these scales for SD. Thus, while the control for SD has reduced the total size of the evaluative dimension it has not eliminated it from the 18 modified scales used here, particularly in the responses of Ss like these psychiatric patients.

However, comparison of the arrays of evaluative factor loadings in Tables 1 and 2 shows that, aside from the four unbalanced scales, only two scales (fragile-rough and easygoing-aggressive) have high loadings in both groups, while one scale (simple-complex) is reversed in direction. This relatively low agreement between the two groups on the evaluative dimension is reflected in an index of factor similarity of only 0.72 between the two evaluative factors (to be discussed further below). It would appear that the scales which are balanced for SD can take on evaluative connotations but may do so differently for different subjects.

### *Factorial correspondence between groups*

Five of the six factors of Tables 1 and 2 have been given tentative names, but aside from the evaluative factors, no effort will be made to describe them. Identification should be deferred until they are cross-validated by further study.



Indices of factor similarity between the factors are reported in Table 3. This matrix of indices approximates the form of an identity matrix, at least for the factors tentatively named 'Evaluation', 'Dominance', 'Liberalism' and 'Expression'. The remaining two factors in each matrix cannot be so clearly matched. What is more, it certainly cannot be said that any of the factors show a high degree of similarity between the two groups. None of these indices of factor similarity can compare with those of 0.9 and higher reported by Kumata & Schramm (1956) or by Suci (1960) for his two English-speaking control groups. Triandis & Osgood (1958) set a criterion of 0.75 for significance of an index of factor similarity. The only index in Table 3 that approaches this value is that between evaluative factors, which include the four unbalanced marker scales.

Table 3. *Indices of factor similarity between factors of Tables 1 and 2\**

Patient group	Non-patient Group					
	Eval.	Dom.	Lib.	Exp.	Tens.	(Unnamed)
Evaluation	0.72	—	—	—	-0.32	—
Dominance	—	0.33	—	—	—	—
Liberalism	—	—	0.55	—	-0.41	—
Expression	—	—	—	0.51	—	—
Tension	0.31	—	—	—	0.51	—
(Unnamed)	—	—	—	—	0.41	0.24

\* Only coefficients above 0.30 are reported.

### *Factorial correspondence between concepts*

The six rotated factor matrices for individual concepts are not reported here, but comparisons between them are presented in Tables 4 and 5. As noted above, four principal axes factors accounting for about 50 per cent of the variance were extracted in each analysis. In each rotated factor matrix, these factors were designated A, B, C and D. Evaluative factors can be identified in each case by the occurrence of high loadings on the four marker scales, but no attempt has been made to identify other factors.

Tables 4 and 5 report indices of factor similarity between the rotated factor matrices of each individual concept. Evaluative factors are indicated by an asterisk. Table 4 reports comparisons between concepts for each group. As predicted, there is little similarity between these factors. Highest indices usually occur between evaluative factors, and, in most cases, it is difficult to match the remaining factors of one concept with those of another. Thus, factors appear to be relatively specific to concepts, with a high degree of concept-scale interaction.

An interesting exception is the comparison between the concepts *patient* and *myself* for the patient group. This four-by-four matrix shows relatively high indices of similarity between three of the factors for each concept, especially an index of 0.80 between evaluative factors. In the non-patient group, these two concepts show nothing like this degree of correspondence.

Table 5 reports comparisons between groups for each of the three concepts.

Here, high factorial correspondence was predicted, since a concept should have essentially the same underlying connotative dimensions for linguistically similar Ss. The results do not bear this out in the case of the concepts *doctor* and *patient*, and for all concepts, indices of factor similarity are much lower than those reported in other studies comparing groups. For the concept *doctor*, even evaluative factors are difficult to equate, suggesting that Ss who are hospitalized psychiatric patients differ greatly from those who are not in the evaluative criteria they apply to such a concept. The patients' description of the concept *doctor* actually yields two factors, orthogonal to each other, that can be identified as evaluative. One of these (Factor 'A') has no counterpart in the non-patient concept *doctor*.

Table 4. *Indices of factor similarity between single concepts*

		<i>Patient group</i>							
		<i>Patient factors</i>				<i>Myself factors</i>			
<i>Doctor factors</i>		A*	B	C	D	A*	B	C	D
A*		0.73	—	—	—	0.68	—	—	0.40
B*		-0.47	—	0.49	—	-0.44	—	0.44	0.39
C		—	—	—	0.31	—	—	—	—
D		0.42	—	—	—	0.48	0.33	—	—
<i>Patient factors</i>									
A*						0.80	—	—	—
B						—	—	—	0.60
C						—	0.32	0.76	0.36
D						—	—	—	—
		<i>Non-patient group</i>							
		<i>Patient factors</i>				<i>Myself factors</i>			
<i>Doctor factors</i>		A*	B	C	D	A	B*	C	D
A		—	0.42	0.40	—	0.38	0.32	0.52	—
B*		0.37	—	—	0.35	—	0.61	—	0.34
C		—	0.47	—	—	—	—	—	—
D		0.37	—	—	—	0.52	—	0.44	—
<i>Patient factors</i>									
A*						0.44	0.60	—	0.42
B						0.54	—	—	—
C						0.43	—	0.31	—
D						0.30	—	0.32	—

(Only coefficients above 0.30 reported.)

\* Evaluative factors.

The concept *myself* shows greater factorial correspondence. Though indices of factor similarity here are still only of moderate magnitude, one (that between Factor 'B' for the patient group and Factor 'A' for the non-patient group) exceeds 0.75. The four-by-four matrix of similarity indices approximates an identity matrix, with only one high coefficient in each row and each column, so that the factor structure of this concept for one group can be matched with that for the other. The magnitude of indices and the form of the matrix are like Table 3, which compared the results of factor analyses based on several concepts.



Table 5. *Indices of factor similarity between groups on single concepts*

Concept: <i>A Doctor</i>				
Patient factors	Non-patient factors			
	A	B*	C	D
A*	—	—	—	—
B*	—	-0.55	-0.38	—
C	—	—	0.38	—
D	-0.41	0.34	—	0.43

Concept: <i>A Patient</i>				
Patient factors	Non-patient factors			
	A*	B	C	D
A*	0.57	—	—	—
B	—	-0.40	—	—
C	—	—	—	—
D	—	—	—	—

Concept: <i>Myself</i>				
Patient factors	Non-patient factors			
	A	B*	C	D
A*	—	0.59	—	—
B	0.78	—	—	—
C	—	—	0.51	—
D	—	—	—	0.56

(Only coefficients above 0.30 reported.)

\* Evaluative factors.

## DISCUSSION

While the control for SD is not complete in these scales, the effort to balance them for favourability seems to have produced an instrument that is more sensitive to subject differences than is the usual form of the semantic differential. Where favourability is uncontrolled, as in semantic differentials used in other studies, a high degree of semantic generality is the usual finding, even when subjects are drawn from culturally and linguistically different populations. Bopp (Osgood *et al.*, 1957) compared schizophrenic and normal subjects and found virtual identity of factor structure. Several cross-cultural studies comparing different or even unrelated languages such as Greek (Triandis & Osgood, 1958), Japanese (Kumata & Schramm, 1956) or Navajo, Hopi and Zuni (Suci, 1960) with English report a high degree of factorial correspondence. Yet in the present study, the responses of two groups of subjects differing in some situational and socio-economic characteristics but speaking the same language and belonging to the same culture, show only moderate factorial correspondence, at best.

This is most apparent for concepts other than the self. It is tempting to speculate that situational differences between these two groups account for the differences found in such concepts as *doctor* and *patient*, which are likely to have quite different meaning for hospitalized psychiatric patients than for non-hospitalized persons. In addition, the *Ss* who are, themselves, patients show relatively good factorial correspondence between the single concepts *patient* and *myself*, while non-patient

Ss do not. The concepts *doctor* and *patient* call for stereotypes, but, with favourability controlled, these stereotypes appear to be based on rather different underlying dimensions in different situations.

The less stereotyped (but more specific) concept *myself*, shows much greater similarity between the two groups in its descriptive dimensions. For both groups, standard deviations of scores in the concept *myself* run slightly higher than in the *doctor* and *patient* concepts. In this case, each subject is describing a different concept and, for most scales, scores are likely to distribute themselves along the entire bipolar scale. The resulting heterogeneity of scale usage more closely approximates the heterogeneity obtained when several concepts are combined in the scale intercorrelations. Greater factorial correspondence between the two subject groups is found under these circumstances. Possibly relative absence of stereotypy (or wider scale score range) allows for the emergence of more 'basic' and common dimensions, of a factor structure which is not distorted by the phenomenon that Osgood (1962) terms 'factorial coalescence' in which factors are correlated and coalesce mathematically, thus distorting the remaining factors in the matrix.

Another way of looking at these results is in terms of the relatively high degree of subject-concept-scale interaction indicated by the findings. Osgood (1962) has recently discussed the relationship between high concept-scale interaction and denotative as opposed to connotative relationships between noun and adjective. It is possible that the particular concepts and scales used in this study are more likely to have denotative than connotative relationships for certain subjects. That is, in a group of hospitalized patients a doctor may always be seen as practical rather than creative, perhaps because of contact with particular doctors, while this adjective-noun relationship may not occur for non-patient subjects nor for other concepts as described by the same patients. It is difficult, however, to see how there could be such a shift in the nature of the adjective-noun relationships of these particular concept-scale combinations as compared to those of other semantic differential studies.

#### SUMMARY

A control for SD response set was introduced into a semantic differential designed for use with person concepts. Eighteen scales were balanced for the SD values of their bipolar adjectives and four additional unbalanced evaluative scales were added as marker items. This 22-scale semantic differential was used by two groups of subjects, 52 hospitalized psychiatric patients and 40 non-patient subjects, to describe three concepts, *a doctor*, *a patient in a hospital* and *myself*.

Responses were factor analysed by groups and by concepts, and factor structures were compared. Examination of evaluative factors indicated that, while only partial control for SD had been attained, the amount of variance attributable to favourability had been reduced, and the evaluative dimension was different for the two groups. All comparisons showed a much lower degree of factor similarity between the two groups of English-speaking subjects than is usually found with the semantic differential, even in cross-cultural studies. Greatest similarity occurred between evaluative factors containing the four unbalanced marker items.

It was concluded that the introduction of a control for SD in the semantic



differential makes the scales more sensitive to subject differences. In this case, situational differences seem to account for lack of correspondence in the factor structures of the two patient-relevant concepts. Some possible explanations for the high degree of subject-concept-scale interaction found here were discussed.

The author wishes to express her gratitude to Charles B. Schultz for his help with machine computation in the factor analyses, and to both Dr Schultz and Dr William S. Kogan for reading and helping to improve the manuscript of this paper.

The six rotated factor matrices may be obtained from the author on request.

# REFERENCES

- BASS, B. M. (1955). Authoritarianism or acquiescence. *J. abnorm. soc. Psychol.* **51**, 616-623.
- BRONFENBRENNER, U. (1958). The study of identification through interpersonal perception. In *Person Perception and Interpersonal Behavior* (eds. Tagiuri, R. & Petrullo, L.), pp. 110-130. Stanford: Stanford Univ. Press.
- COUCH, A. & KENISTRON, K. (1960). Yeasayers and naysayers: agreeing response set as a personality variable. *J. abnorm. soc. Psychol.* **60**, 151-174.
- COWEN, E. L. & TONGAS, P. N. (1959). The social desirability of trait descriptive terms: applications to a self-concept inventory. *J. consult. Psychol.* **23**, 361-365.
- CRONBACH, L. J. (1946). Response sets and test validity. *Educ. psychol. Meas.* **6**, 475-494.
- CRONBACH, L. J. (1950). Further evidence on response sets and test design. *Educ. psychol. Meas.* **10**, 3-31.
- EDWARDS, A. L. (1953). The relationship between judged desirability of a trait and the probability that the trait will be endorsed. *J. appl. Psychol.* **37**, 90-93.
- EDWARDS, A. L. (1955). Social desirability and Q sorts. *J. consult. Psychol.* **19**, 462.
- EDWARDS, A. L. (1957). *The Social Desirability Variable in Personality Assessment and Research*. New York: Dryden Press.
- HELPER, M. M. (1958). Parental evaluations of children and children's self-evaluations. *J. abnorm. soc. Psychol.* **56**, 190-194.
- JACKSON, D. N. & MESSICK, S. J. (1957). A note on 'ethnocentrism' and acquiescent response sets. *J. abnorm. soc. Psychol.* **54**, 132-133.
- KUMATA, H. & SCHRAMM, W. (1956). A pilot study of cross-cultural meaning. *Publ. Opin. Quart.* **20**, 229-237.
- LAZOWICK, L. M. (1955). On the nature of identification. *J. abnorm. soc. Psychol.* **51**, 175-183.
- MARLOWE, D. & CROWNE, D. P. (1961). Social desirability and response to perceived situational demands. *J. consult. Psychol.* **25**, 109-115.
- NUNNALLY, J. & KITROSS, J. M. (1958). Public attitudes toward mental health professions. *Amer. Psychologist*, **13**, 589-594.
- OSGOOD, C. E. (1952). The nature and measurement of meaning. *Psychol. Bull.* **49**, 197-237.
- OSGOOD, C. E. (1962). Studies on the generality of affective meaning systems. *Amer. Psychologist*, **17**, 10-28.
- OSGOOD, C. E. & LURIA, Z. (1954). A blind analysis of a case of multiple personality using the semantic differential. *J. abnorm. soc. Psychol.* **49**, 579-591.
- OSGOOD, C. E. & SUCI, G. J. (1952). A measure of relation determined by both mean difference and profile information. *Psychol. Bull.* **49**, 251-262.
- OSGOOD, C. E., SUCI, G. J. & TANNENBAUM, P. H. (1957). *The Measurement of Meaning*. Urbana: Univ. of Illinois Press.
- SUCI, G. J. (1960). A comparison of semantic structures in American South-west culture groups. *J. abnorm. soc. Psychol.* **61**, 25-30.
- TRIANDIS, H. C. & OSGOOD, C. E. (1958). A comparative factorial analysis of semantic structures in monolingual Greek and American college students. *J. abnorm. soc. Psychol.* **57**, 187-196.

*Manuscript received 25 June 1962*

*Revised manuscript received 16 April 1963*

## Attitude Changes Associated with Training in Human Relations

By PETER B. SMITH\*

*Industrial Management Division, The University of Leeds*

Members of training groups in human relations change their attitudes toward social behaviour in a way that members of other groups do not. Those trained showed a convergence toward median scores on scales measuring their attitudes toward power and close personal relationships. These attitudes were found to be related to perceptions of their actual behaviour by other group members. Perceptions of the behaviour of those scoring median on the attitude scales were generally consonant with current descriptions of 'effective' behaviour.

### INTRODUCTION

Training in human relations by non-directive methods has now been under way in the United States for more than 15 years. Despite the pioneering work of Bion (1961) more than 20 years ago in London, only in the last 5 years have such methods been widely used in Britain. The phrase 'training in human relations by non-directive methods' is used here to mean any method which relies primarily on the trainee achieving insights into the nature of social behaviour by discovery rather than by formal instruction. The intentions of those developing non-directive training have been to set up conditions which facilitate such learning by discovery. This has normally meant bringing together a group of people for the purpose of studying the behaviour of their own group, with the aid of a staff member, often known as the trainer. The trainer's task is to indicate by occasional interventions, in the form of questions or tentative interpretations, the areas of behaviour which he regards as providing useful data for study. Such trainer interventions often do not focus directly on the tasks which the groups may set themselves, but on the processes by which the goals of the group are achieved. The development of these methods in Britain has been surveyed by Crichton (1962).

The assessment of the insights which are achieved by those undergoing training poses methodological problems which few studies have yet overcome. Miles (1959, 1960) has provided a concise summary of contemporary training aims in terms of increases in:

'*Sensitivity*: the ability to perceive what is actually going on in a social situation (including both behavioural events and inferred feelings of other persons).

*Diagnostic ability*: the skill of assessing on-going social situations in a way that

\* The data in this paper were collected while the author was a research student of the University of Cambridge. I am grateful to Professor F. J. Willett for his supervision of the project, to the D.S.I.R. for support, and to the organizers, trainers and members of the various courses. The courses studied were organized by Cambridge and Leeds Universities, Birmingham College of Advanced Technology and the Simon (Engineering) Ltd. group of companies.



enables effective action; the employment of appropriate explanatory categories to understand reasons for presented interaction.

*Action skill*: the ability to intervene effectively in on-going situations in such a way as to maximize personal and group effectiveness and satisfaction. . . .'

The most direct test of increases in these three variables would be a study of the job behaviour of those trained, since a crucial component of training must be that the insights achieved are transferable by the individual into his normal life-situation. Miles (1960) showed that the job behaviour of 34 schoolteachers was perceived to change after training, both by the teachers themselves and by their associates. The teachers perceived themselves in ways which Miles classified as 'more sensitive to the needs of others' and 'showing greater action skill', while the associates most often saw the teachers as showing improved leadership skills and communicating more openly. These changes were not found among the controls. In a study by Harrison, reported in Argyris (1962), changes were examined in the ways in which those who had been trained perceived associates who had not. It was found that descriptions by at least some managers of their untrained associates showed an increased use of 'interpersonal-emotional' terms. The controls described their associates uniformly in terms of 'rational-intellective' concepts. The trained managers were thus utilizing data about their associates which the controls were discarding as unimportant. This change would be most closely related to increases in Miles' variable sensitivity. These and other studies of training outcomes are discussed more fully in Smith (1962b).

The present study examines changes in the attitudes toward social behaviour of those trained; such changes may *imply* increases in Miles' three variables. The attitudes considered were those toward control behaviour, which is concerned with power, and affection behaviour, which is concerned with close personal relationships. Control and affection behaviours are among the most important in the interaction of non-directive training groups. Issues relating to control behaviour often focus on the role of the trainer who, by refusing to lead in the expected manner, sets the group a leadership problem which it must solve. Likewise, issues relating to affection behaviour usually arise as concerns about how much each individual shall reveal of his personal feelings. The training group is a situation which arouses mild anxiety in many people. Faced with ambiguity, each person soon shows his characteristic patterns of origination of control and affection behaviours.

Attitudes toward control and affection behaviours were measured on four Guttman scales derived from Schutz's (1958) FIRO. For each behaviour one scale (the 'expected' scale) measures how often the respondent expects to show the behaviour himself, and the second scale (the 'wanted' scale) measures how often the respondent wants others to show the behaviour. If the simple postulate is established that we seek to repeat those patterns of behaviour that we find rewarding, both scales can be considered as yielding data as to what the respondent finds more rather than less rewarding. A respondent with a high 'expected' score and a low 'wanted' score will be most highly rewarded when his behaviour is more active than that of others in his group. Conversely, a respondent with a low 'expected'



score and a high 'wanted' score will be most highly rewarded when he can behave more passively than most members of his group. In this paper the difference between a man's 'expected' and 'wanted' scores will be referred to as his reward score. The use of the reward score will be to predict whether an individual prefers control behaviour to affection behaviour. As such it clearly measures only the *relative* rewards derivable by the individual from these two behaviours; there is not usually any reason to expect reward scores to relate to an individual's *overall* reward levels, for example whether he is satisfied or dissatisfied with his membership of a group. This distinction is a familiar one in economics, and has recently been discussed at length in Homans (1961) and Thibaut & Kelley (1959).

Reward scores were divided as nearly as possible into equal thirds described as positive, zero and negative. A positive reward score is one for which the 'expected' scores exceeds the 'wanted' score.

Schutz suggests that those with positive or negative reward scores tend to respond to anxiety-provoking situations in a fixed way, whereas those with zero reward scores are less anxious and therefore more able to respond to any situation adaptively. For example, in the case of control behaviour, the negative reward scorer tends to avoid or withdraw from situations in which he must exercise control, so as to minimize his anxieties about controlling people; the positive reward scorer exercises compulsive control wherever he can, in order to reassure anxieties that he is basically incapable of controlling others; but the zero reward scorer is able both to control and to be controlled, as the situation demands, without undue anxiety.

The changes resulting from non-directive training will vary with the personality of each individual. Those with positive reward scores may be expected to learn of some of the unintended effects of their behaviour, and to become more skilled in regulating their pattern of interventions in accord with the needs of the situation. Those with negative reward scores may learn of the greater rewards to be derived from active intervention. A convergence toward the median may therefore be predicted. This expectation may be compared with Berlew's (1960) findings relating sensitivity to motive strength. Using projective tests, he found that ability to predict other people's perceptions of their own behaviour was greatest among those who scored median on tests of need-power and need-affiliation. If projective measures of need-power and need-affiliation can be considered as equivalent to Guttman scale measures of attitudes toward control and affection, Berlew's findings indicate maximum sensitivity among median scorers.

The mechanism whereby such convergence toward the median might occur can be readily envisaged. One of the processes encouraged by trainers is the 'giving and receiving of feedback'. This is a process whereby group members tell each other what impact their behaviour has had on them. Training group members thus obtain a much fuller picture of the effects of their behaviour on others than they do in everyday life. The permissive atmosphere of a training group provides the member with an opportunity to try out and learn new roles or methods of responding to familiar situations.

The prediction of appreciable changes in attitudes after training poses the question of whether the attitudes are related to actual behaviour. Since the FIRO



scales used take the form of action-oriented statements (e.g. 'I let other people decide what to do'), a positive relationship may be foreseen. Borg (1960), Sapolsky (1960) and Smith (1962a) have reported significant relationships between FIRO scores and independent measures of actual behaviour, such as amount of verbal participation.

### METHOD

All subjects were either managers on training courses or undergraduates studying management at various universities. All but two were men. The managers were predominantly graduates of between 30 and 40. The 108 experimental subjects were members of eleven non-directive training groups, four of undergraduates and seven of managers. The 44 control subjects were members of six discussion groups, one of undergraduates and five of managers. These discussions were led by a staff member, acting in a more or less directive role, and considered aspects of social psychology relevant to management problems. Manager groups mostly met intensively for a few weeks while undergraduate groups lasted through the academic year.

The attitude measure used was Schutz's (1958) FIRO questionnaire, which consisted originally of six Guttman scales describing the respondent's perception of how often he performs certain behaviours and how often he likes others to do so. The four scales used in this study were slightly revised for British usage, and were those describing attitudes toward control and affection behaviours.

Subjects completed the FIRO questionnaire near the beginning and again near the end of their course. At the same time all but three groups made checklist nominations of the behaviours which they had perceived in other members of their group. These nominations were used to test the validity of the FIRO reward scores. The checklist comprised six control behaviours, six affection behaviours and three behaviours which were not classifiable as control or affection behaviours:

#### *Control behaviours*

- Striving for individual recognition
- Discussing usefulness of meetings
- Dividing up the task
- Making rules or laying down procedures
- Attempting to dominate or to control
- Rebelling or obstructing the group

#### *Affection behaviours*

- Laughing and showing happiness
- Attempting to preserve group unity
- Reconciling antagonisms
- Attempting to draw people in
- Showing close friendship
- Discussing close personal details

#### *Other behaviours*

- Withdrawing from group activity
- Disliking or rejecting others
- Submitting to others' wishes

### RESULTS

The coefficients of reproducibility of the four Guttman scales of FIRO are shown in Table 1. Highly significant test-retest reliability was found for both control and affection reward scores. When positive, zero and negative scores were compared, using the scores obtained early and late during each group's duration, chi-square

for control was 53.3, and for affection 39.4 (Smith, 1962a). Both these values of chi-square are significant for 4 degrees of freedom at  $p < 0.001$ .

While the reward scores are therefore reliable within gross limits, this does not necessarily mean that no changes occurred in the attitudes of those trained. The predicted changes are:

- (i) a fall in positive reward scores, and
- (ii) a rise in negative reward scores.

Table 1. *Coefficients of reproducibility of FIRO scales*

FIRO scale	Coefficient
Expressed control	0.91
Wanted control	0.93
Expressed affection	0.96
Wanted affection	0.94

Table 2 compares the changes in non-directive training groups and the control discussion groups.

In the experimental groups, 91 scores changed in the predicted direction and 32 in the reverse direction. This is highly significant ( $p < 0.001$ : 1-tail binomial test).

Table 2. *Number of subjects showing rise or fall in reward scores*

Subjects	Initial reward scores								
	Positive			Zero			Negative		
	+	=	-	+	=	-	+	=	-
Experimentals									
Control	5	6	29	11	5	11	20	6	15
Affection	4	7	12	11	5	12	30	9	8
Controls									
Control	8	3	11	2	2	5	7	3	3
Affection	3	2	8	5	2	6	8	2	6

In the control groups, 36 scores changed in the direction predicted for the experimental groups and 24 in the reverse direction. This change is not significant. When the change found in the experimental group is compared with that in the control group, the difference is again significant ( $p < 0.001$ ). The most marked changes in the experimental groups were the fall in positive control reward scores and the rise in negative affection reward scores. The method of analysis adopted can give no indication of the magnitude of the changes found, but it does show that they occurred frequently.

The division of control and affection reward scores into equal thirds means that the subjects can be subdivided into nine separate types. However it is not expected that the behaviour of each of the nine types will be differentially perceived by others. For example, subjects with negative reward scores will be predominantly



inactive, making it difficult for others to perceive control or affection behaviours: it is predicted that they will be perceived as showing 'neither control nor affection' behaviours. The nine types have been combined into the five classes for which different predictions may be made. These are:

1. Control reward score positive and affection reward score negative.
2. Affection reward score positive and control reward score negative.
3. Both reward scores positive, or one positive and the other zero.
4. Both reward scores zero.
5. Both reward scores negative, or one negative and the other zero.

The general predictions were made that positive reward scores would be associated with frequently occurring perceptions of control and affection behaviours, while zero reward scores would be associated with a chance occurrence of perceptions of the behaviours, and negative reward scores with a significant lack of perceptions of the behaviours. The specific predictions derivable for each class are shown in Table 3.

The occurrence of each of the 15 behaviours on the checklist was analysed separately. The degree to which each of the five classes was characterized by the various behaviours was established by consideration of the occurrence of checklist nominations in the class, relative to their occurrence in the total sample. Ratios were computed for each behaviour and every class, as follows:

$$\text{Ratio} = \frac{\text{Actual No. of nominations in class}}{\text{Expected No. of nominations in class}}$$

where expected nominations = Total nominations for all classes

X the proportion of total subjects in the class

Table 3 tests whether the most and least characteristic behaviours of members of each class were those that had been predicted from FIRO reward scores. The ratios were rank ordered and the predictions tested by the Mann-Whitney U-test.

Table 3. *Relationship between behaviour predicted from FIRO reward scores and behaviour assessments by other group members*

Class	Prediction	<i>p</i>
1	C will predominate	< 0.02
2	A will predominate	< 0.10
3	C and A will predominate	< 0.01
4	No predominant behaviour	C < N, A n.s. A < C, N n.s. N < C, A n.s.
5	N will predominate	n.s.

Key: C—Control behaviour

A—Affection behaviour

N—Behaviours which are neither control nor affection.

The classes are defined above.

The predictions are significantly upheld for Classes 1, 3 and 4 and approach significance for Class 2. Class 5 contains those group members whose behaviour is least easily perceived, so that the failure of the prediction is not unexpected.

This paper has outlined some of the stated aims of those active in non-directive training. It has also given evidence as to the changes actually found in some training groups. If the training is successful, there should be a close correspondence between training aims and actual changes. The ratios calculated show which of the behaviours on the checklist were perceived as most characteristic of group members of the various personality types derived from the FIRO reward scores. Since those trained tend toward zero reward scores, those who already have zero reward scores should come closest to exemplifying Miles' three training aims—'sensitivity', 'diagnostic ability' and 'action skill', while those with positive or negative scores should be less well placed.

Table 4 shows the distribution between the five classes of the behaviours on the checklist. For each behaviour a value of chi-square was computed to test the randomness of the occurrence of the behaviours. The table includes only those behaviours which were perceived differentially among members of the different classes ( $p < 0.05$ ).

The table shows marked differences between the behaviours characterizing the different classes, and also in the degree to which the classes were characterized by distinct behaviours. Members of Class 1 were characterized by self-oriented high activity control behaviours and a minimum of affection behaviour. Members of

Table 4: *Behaviour of the different classes of training group members*

Behaviours	Class					Chi-square
	1	2	3	4	5	
Control						
Striving for individual recognition	1.68	0.73	1.28	0.51	0.64	36.86
Discussing usefulness of meetings	1.12	0.74	1.17	1.24	0.85	12.94
Dividing up the task	0.89	0.52	1.06	2.25	0.95	11.27
Attempting to dominate or to control	1.57	0.56	1.49	0.36	0.58	38.00
Affection						
Laughing and showing happiness	1.02	0.79	1.25	0.98	0.83	14.36
Reconciling antagonisms	0.40	1.10	1.14	1.86	0.94	10.27
Showing close friendship	0.37	0.96	1.58	0.64	0.78	17.68
Neither						
Withdrawing from group activity	0.82	0.86	0.61	0.98	1.42	22.30

The classes are defined on p. 110. Ratios are obtained by dividing actual numbers of checklist nominations by the number to be expected on the basis of nominations in the total population. The table includes only those behaviours whose distribution between the classes differed from randomness at  $p < 0.05$ . The values of chi-square are shown on the right.

Class 2 were characterized by the absence of control behaviours rather than a high incidence of affection behaviours. Members of Class 3 showed predominantly high activity behaviours, both control and affection, many of which were self-oriented. Conversely, Class 5 showed a lack of high activity control behaviours and typically withdrew from group activity. The most frequent behaviours in Classes 1, 2, 3 and



5 were thus high activity self-oriented behaviours or else withdrawal from group activity.

However, it is the behaviour of Class 4 members which is of particular interest because the class is made up of zero reward scorers, towards which other members of training groups have been shown to change their attitudes. Members of Class 4 were the most clearly characterized of any class, obtaining both the highest and lowest values of the ratio. The characteristic behaviours were those that take into account not only the needs of the individual but also those of the group. In contrast to the infrequent self-oriented behaviours, the Class 4 member was often perceived as 'dividing up the task', 'reconciling antagonisms' and 'discussing usefulness of meetings'.

### DISCUSSION

The training group in human relations is a potent agent for change of attitudes about social behaviour. The effects of the individual's learning about the ways in which others perceive his behaviours may include marked changes in his habitual behaviours and in the way others come to perceive these behaviours. The evidence cited in support of these statements in this paper gives only general indications of their probable truth. Human relations training is currently conducted with many different emphases and no evidence is yet forthcoming as to whether these different emphases affect the training outcomes. The two Leeds training groups were conducted as 'sensitivity training', which usually means that the trainer indicates a greater preference for discussing participants' feelings rather than the development of group roles. On the other hand the six Simon (Engineering) training groups had trainers who behaved in more varied ways and the groups participated also in more formal lectures and exercises in social psychology. The three Cambridge training groups showed an intermediate emphasis.

Burke & Bennis (1961) reported increases in similarity of perceptions of individuals by self and others in training groups. Such a study gives no evidence as to increases in *sensitivity*, since perceptions late in the life of a training group are based on much greater information than earlier perceptions. The present study, in contrast, shows that attitudes changed toward the median, and that median scorers were perceived in ways that have been described; but early perceptions were not separated from late ones. Thus the changes found could not be artifacts derived from the possibility that training group members may come to know each other much better than the control group members.

Early writers in the field of human relations often stressed the need to reduce or eliminate social conflict. Their critics (e.g. Whyte (1956), McNair (1957)) tended to feel that this emphasis represented an attack on individuality, a too-ready advocacy of submission to group pressures in the interests of smooth social working. Recent writers on human relations training have laid much more emphasis on the need to recognize and bring into the open substantive conflicts (e.g. Argyris, 1962). *Effective group behaviour is seen as that which is based on a realistic knowledge of differences of viewpoint as well as of similarities. Parallel thinking is implicit in Miles' definition of 'action skill' cited earlier, where he states that interventions*

should maximize both personal and group satisfactions. Where these are not reconciled there will be conflict. This conception of effective behaviour is well illustrated by the present findings. The zero reward scorer is perceived as showing close friendship as infrequently as the self-oriented control behaviours. Indeed there is a tendency for him also to be perceived as 'rebellious or obstructing the group', but chi-square for this behaviour does not quite achieve the required significance level.

The changes in attitudes found can be seen as a response to a situation in which customary behaviours proved to be inappropriate. Whether the changes persist will depend primarily on the reactions of the trained man's associates in the weeks after his return from the course, and on the degree of continuing support which he receives from others who participated in the training.

### REFERENCES

- ARGYRIS, C. (1962). *Interpersonal Competence and Organizational Effectiveness*. London: Tavistock.
- BERLEW, D. E. (1961). Interpersonal sensitivity and motive strength. *J. abnorm. soc. Psychol.* **63**, 390-394.
- BION, W. R. (1961). *Experiences in Groups*. London: Tavistock.
- BORG, W. R. (1960). The prediction of role behaviour in small groups from personality variables. *J. abnorm. soc. Psychol.* **60**, 112-117.
- BURKE, R. L. & BENNIS, W. G. (1961). Changes in perceptions of self and others during human relations training. *Hum. Relat.* **14**, 165-182.
- CRICHTON, ANNE (1962). Personnel Management and Working Groups. *Institute of Personnel Management, Occasional Papers*, **18**.
- HOMANS, G. C. (1961). *Social Behaviour: Its elementary forms*. New York: Harcourt, Brace.
- MCNAIR, M. (1957). What price human relations? *Harvard Bus. Rev.* **35**.
- MILES, M. B. (1959). *Learning to Work in Groups*. New York: Teachers' College, Bureau of Publications.
- MILES, M. B. (1960). Human Relations training: processes and outcomes. *J. couns. Psychol.* **7**, 301-306.
- SAPOLSKY, A. (1960). Effect of interpersonal relationships upon verbal conditioning. *J. abnorm. soc. Psychol.* **40**, 241-246.
- SCHUTZ, W. C. (1958). *FIRO: A Three-dimensional Theory of Interpersonal behaviour*. New York: Rinehart.
- SMITH, P. B. (1962a). Role differentiation in small social groups. Cambridge: Unpublished Ph.D. thesis.
- SMITH, P. B. (1962b). A survey of research into attitude and behaviour changes resulting from 'human relations' training. *Working Papers in Industrial Management*, **1**. Industrial Management Division, University of Leeds.
- THIBAUT, J. W. & KELLEY, H. H. (1959). *The Social Psychology of Groups*. New York: Wiley.
- WHYTE, W. H. (1956). *The Organisation Man*. New York: Simon & Schuster.

*Manuscript received 8 February 1963*

*Revised manuscript received 27 April 1963*



## The Effect of Attitudinal Factors on the Relationship Between Conditioning and Personality

By IHSAN AL-ISSA

*Institute of Psychiatry (Maudsley Hospital), London University*

The aim of this study was to investigate the effect of attitudinal factors on the relationship between eye-blink conditioning and personality measures of extraversion, neuroticism and anxiety. The subject's attitudes were experimentally manipulated by (1) varying the instructions; (2) introducing a reaction time task. Under these conditions, an attempt was made to test the prediction made by Eysenck that eye-blink conditioning is negatively correlated with extraversion as measured by the Maudsley Personality Inventory. Another prediction tested concerns the Spence-Hull hypothesis that there is positive correlation between eye-blink conditioning and personality tests of neuroticism and anxiety. Results show that the relationship between eye-blink conditioning and personality measures is inconsistent and varies with experimental conditions. This inconsistency was due in part to the new scoring method employed in the present study. Results are discussed within the framework of previous studies relating conditioning to personality measures.

### 1. INTRODUCTION

In the past few years, studies relating conditioning to individual differences have been stimulated primarily by Spence and Taylor, and by Eysenck and his associates. These studies were extensively reviewed by Eysenck (1962) and Martin (1961). Consideration of all experiments to date indicates that the data are inconsistent, even contradictory, and serves to demonstrate that experimentation on these lines gives rise to artifacts through lack of control of relevant factors. The method of scoring the conditioning records, the selection of samples, the presence or absence of a ready signal preceding each conditioning trial, and the instructions given to subjects, are some of the many factors involved in the experimental situation.

The present study was designed to investigate the effect of different instructions on the relationship between eye-blink conditioning and personality. Under these conditions it was hoped to test the predictions made by Eysenck that eye-blink conditioning is negatively correlated with extraversion. Another prediction tested concerns the Spence-Hull hypothesis that there is positive correlation between eye-blink conditioning and scores on personality tests of neuroticism and anxiety.

### 2. METHOD

#### *Apparatus*

A detailed description of the eye-blink conditioning apparatus is given by Al-Issa (1962) and Eysenck (1963). The conditioned stimulus (CS) was a pure tone with a frequency of 1110 c.p.s. and an intensity of 60 dB above the subject's hearing threshold, lasting for 500 m-sec. The unconditioned stimulus (UCS) was an air puff at a pressure of 3 psi (pounds per square inch) and of 60 m-sec duration, ejected at a distance of approximately 1 cm from the right eye.

### *Subjects*

Subjects were 90 male adolescent industrial apprentices at the Ford Motor Company. The ages of the main three groups ranged from 15-17 years. No subject had previous experience with the eye-blink conditioning test.

### *Procedure*

Three groups of 30 subjects were used to test the conditioning and extinction of eye-blink responses. One group, 'the no-task' group, was instructed in such a way that the emphasis was put on the reaction of the subjects to the air puff (UCS). Members of the second group, the 'active-task' group, were given a reaction time task (key switch) and were instructed to react to the tone (CS) as quickly as they could. The air puff (UCS) was considered in this situation as an aversive stimulus to punish subjects for their slow reaction time. Instructions to this group mainly stressed the reaction time task and were intended to deceive subjects as to the main purpose of the experiment. If a subject belonged to the third group, the 'passive-task' group, he was first asked to press a key switch and let his finger rest passively on it throughout the test of conditioning. Then he was given the same instructions as the 'no-task' group. During the test for extinction, the 'active-task' group and the 'passive-task' group were split into two halves—15 subjects in each. While one half of each group was allowed to continue the task (key switch) without interruption, instructions were given to the other two halves to stop the task. The two subgroups who used no task in the test for extinction were called the 'active' group and the 'passive' group. All groups were given 20 and 10 trials for the tests of conditioning and extinction respectively.

Subjects were required to complete the Maudsley Personality Inventory (Eysenck, 1959; Jensen, 1958) and the Taylor Manifest Anxiety Scale (Taylor, 1951). Some of the personality scores were discarded when subjects did not complete all items in the questionnaire.

### *Scoring*

In a recent investigation by Spence & Ross (1959) it was demonstrated that when two judges independently classified eye-blink responses within the duration of the UCS (500 m-sec), these responses tended to have distinctive latency distributions. Thus, responses which occurred within a latency of 0-200 m-sec were called alpha eye-blinks; responses which occurred within a latency of 200-300 m-sec were called voluntary eye-blinks, and responses which occurred within a latency of 300-500 m-sec were called conditioned eye-blinks. This latency criterion was employed to score the eye-blink records in the present investigation.

## 3. RESULTS

### *Eye-blink conditioning and extraversion*

In the test for conditioning, it will be seen from Table 1 that, with the exception of the correlation relating alpha eye-blink responses to extraversion which is significant at the 0.05 level, all the remaining eight correlations are non-significant. Similarly, 15 correlations reported in Table 1 between extraversion and eye-blink responses in the test for extinction are not significantly different from zero. Since almost all the correlations between eye-blink responses and extraversion are non-significant (20 out of 21) and point to the positive and negative directions, it is safe to conclude that the present results give no substantial evidence to support the hypothesis that eye-blink conditioning is negatively related to extraversion.

### *Eye-blink conditioning and neuroticism*

An inspection of Table 2 shows that in the test for conditioning three out of nine correlations are significant. On the one hand, conditioned eye-blink responses given by the 'passive-task' group are negatively correlated with the neuroticism scores and significant at the 0.01 level. On the other hand, alpha and voluntary



Table 1. *Correlation coefficients between extraversion (E) as measured by the Maudsley Personality Inventory (MPI) and the frequency of different kinds of eye-blink responses in the tests of conditioning and extinction*

	Conditioning	Extinction
I. 'No-task' group (N = 30)		
Alpha × E	0.09	0.14
Voluntary × E	0.10	0.23
Conditioned × E	0.20	0.10
II. 'Active-task' group		
Alpha × E	-0.42* (N = 28)	-0.01 (N = 13)
Voluntary × E	-0.09 (N = 28)	0.12 (N = 13)
Conditioned × E	-0.30 (N = 28)	-0.49 (N = 13)
III. 'Active' group		
Alpha × E	—	0.32 (N = 15)
Voluntary × E	—	0.09 (N = 15)
Conditioned × E	—	0.12 (N = 15)
IV. 'Passive-task' group		
Alpha × E	0.32 (N = 29)	0.10 (N = 14)
Voluntary × E	0.01 (N = 29)	0.10 (N = 14)
Conditioned × E	0.00 (N = 29)	-0.02 (N = 14)
V. 'Passive' group		
Alpha × E	—	0.26 (N = 15)
Voluntary × E	—	0.04 (N = 15)
Conditioned × E	—	0.42 (N = 15)

\* Significant at 0.05 level.

Table 2. *Correlations coefficients between neuroticism (Neu) as measured by the Maudsley Personality Inventory (MPI) and the frequency of different kinds of eye-blink responses in the tests of conditioning (C) and extinction (E)*

	Conditioning	Extinction
I. 'No-task' group (N = 30)		
Alpha × Neu	-0.09	0.10
Voluntary × Neu	0.12	-0.13
Conditioned × Neu	-0.06	0.14
II. 'Active-task' group		
Alpha × Neu	0.42* (N = 28)	0.49 (N = 13)
Voluntary × Neu	0.45* (N = 28)	0.45 (N = 13)
Conditioned × Neu	0.31 (N = 28)	-0.18 (N = 13)
III. 'Active' group		
Alpha × Neu	—	-0.05 (N = 15)
Voluntary × Neu	—	0.25 (N = 15)
Conditioned × Neu	—	-0.13 (N = 15)
IV. 'Passive-task' group		
Alpha × Neu	-0.08 (N = 29)	0.31 (N = 14)
Voluntary × Neu	0.11 (N = 29)	0.10 (N = 14)
Conditioned × Neu	-0.50† (N = 29)	-0.14 (N = 14)
V. 'Passive' group		
Alpha × Neu	—	-0.04 (N = 15)
Voluntary × Neu	—	0.23 (N = 15)
Conditioned × Neu	—	-0.36 (N = 15)

\* Significant at 0.05 level

† Significant at 0.01 level

eye-blink responses given by the 'active-task' group are positively correlated with neuroticism and significant at the 0.05 level. When correlations concerned with the experimental groups in the test for extinction are considered in Table 2, no consistent or significant results are obtained. In general, results give no conclusive evidence as regards the relationship between eye-blink conditioning and neuroticism.

### *Eye-blink conditioning and manifest anxiety*

In the test for conditioning, it will be seen from Table 3 that two positive correlations concerning the relationship between alpha and conditioned eye-blinks

Table 3. *Correlation coefficients between the Manifest Anxiety Scale (MAS) and the frequency of different kinds of eye-blink responses in the tests of conditioning (C) and extinction (E)*

	Conditioning	Extinction
I. 'No-task' group (N = 12)		
Alpha × MAS	-0.02	0.35
Voluntary × MAS	-0.13	0.24
Conditioned × MAS	-0.05	0.18
II. 'Active-task' group		
Alpha × MAS	0.49* (N = 19)	0.00 (N = 5)
Voluntary × MAS	0.23 (N = 19)	0.82 (N = 5)
Conditioned × MAS	0.73† (N = 19)	-0.91* (N = 5)
III. 'Active' group		
Alpha × MAS	—	-0.16 (N = 14)
Voluntary × MAS	—	0.03 (N = 14)
Conditioned × MAS	—	-0.19 (N = 14)
IV. 'Passive-task' group		
Alpha × MAS	-0.15 (N = 30)	0.72† (N = 15)
Voluntary × MAS	-0.09 (N = 30)	0.08 (N = 15)
Conditioned × MAS	-0.43* (N = 30)	0.16 (N = 15)
V. 'Passive' group		
Alpha × MAS	—	-0.35 (N = 15)
Voluntary × MAS	—	-0.21 (N = 15)
Conditioned × MAS	—	-0.79* (N = 15)

\* Significant at 0.05 level

† Significant at 0.01 level

given by the 'active-task' group and the manifest anxiety scores are in accordance with prediction and significant at the 0.05 and 0.01 level. However, the correlation between conditioned eye-blink responses given by the 'passive-task' group and the manifest anxiety scores is negative and significant at the 0.05 level. Moreover, the six remaining correlations in the test for conditioning are non-significant.

In the test for extinction, Table 3 shows that out of 15 correlations between eye-blink responses and the manifest anxiety scores, four are significant at or beyond the 0.05 level. Again, these correlations are inconsistent and point to the positive and negative directions. Thus, the evidence presented in Table 3 is inconclusive as regards the relationship between manifest anxiety and eye-blink conditioning.



## 4. DISCUSSION

The results reported in this study add to the confusion which is characteristic of previous studies relating conditioning to the personality measures of extraversion, neuroticism and anxiety. They show, however, that contradictory results in this field of research may be due at least in part to the method employed for scoring eye-blink records, and the attitude of *Ss* to, and their knowledge of, the nature of the conditioning experiment. Results reported in Table 1 show that only when alpha eye-blinks given by the 'active-task' group are correlated with extraversion, is the correlation significant at the 0.05 level. The other two correlations between voluntary and conditioned eye-blink responses and extraversion scores for the same group are not significantly different from zero; they are similar to the ones reported by O'Connor & Rawnsley (1959), Willett (1960) and Field (1961). It is thus obvious that the method employed in assigning responses to the alpha, voluntary or conditioned category may be regarded as a deciding factor in the amount of relationship between eye-blinks and extraversion.

Different instructions to *Ss* are as important as the method of scoring in their effect on the relationship between eye-blink responses and extraversion. Although non-significant, such a relationship referring to the 'no-task' and 'passive-task' groups points in the positive instead of the negative direction. As instructions to these two groups were meant to stress the conditioning stimuli and differed from those given to the 'active-task' group, conflicting results indicate the effect of changes in the attitude of the experimental groups. It is noteworthy that other experiments relating conditioning to extraversion have paid no attention to this factor. The results of these experiments (O'Connor & Rawnsley, 1959; Willett, 1960; Field, 1961) are, nevertheless, similar to those of the 'active-task' group whose instructions emphasized the reaction-time task and decreased the 'awareness' of *Ss* as to the nature and purpose of the conditioning experiment. This similarity may be due to the unwitting suggestion given by the experimenters, i.e. by asking the *Ss* to respond naturally to the conditioning stimuli and not to inhibit their eye-blinks, or it may be due to self-initiated instructions.

Again, results presented in Tables 2 and 3 may emphasize similar experimental conditions and factors that are responsible for the inconsistency of the previously reported correlations between eye-blink conditioning and anxiety or neuroticism (Franks, 1954, 1956, 1957; Spence & Taylor, 1953; Willett, 1960; Field, 1961). However, *Ss* general level of performance in eye-blink conditioning may be taken into consideration to evaluate these contradictory correlations. In comparison with other experimental groups, the rapid rate of conditioning and the slow rate of extinction which was demonstrated by the 'active-task' group (Al-Issa, 1962) may account for the tendency of its eye-blink responses to correlate positively with neuroticism and anxiety. This interpretation is in line with previous studies showing that the relationship between eye-blink conditioning and anxiety is dependent to some extent on factors that facilitate or interfere with conditioning such as the presence or absence of a ready signal preceding each conditioning trial (Spence & Taylor, 1953; King *et al.*, 1961; Klinger & Prokasy, 1962). Among many other possible effective factors are differences in the average age of subjects (Braun &

Geiselsart, 1959) and the strength of the air-puff (UCS) employed in the conditioning experiment (Spence & Taylor, 1951).

An interesting question that arises from studies relating personality to conditioning is how far they throw light on theories postulating individual differences in the cortical processes of excitation as investigated by Eysenck (1955, 1957), or individual differences in the level of drive as introduced by Hull (1943) and developed by Spence and his associates (Spence & Taylor, 1951, 1953; Spence & Farber, 1953). The present study draws the attention of research workers to the multiplicity of factors which might affect their prediction concerning the relationship between eye-blink conditioning and different measures of personality. It is therefore most probable that what is measured by the MPI or the MAS 'is an interacting variable and has an effect upon conditioning only when it is combined with the proper values of other variables not explicitly manipulated by E' (King *et al.*, 1961). Unless the effects of many of these variables are elucidated experimentally, conclusions drawn from the present studies about general psychological theories such as those postulated by Eysenck or Hull and Spence would be very premature.

The writer is grateful to Professor H. J. Eysenck under whose supervision this study was conducted.

#### REFERENCES

- AL-ISSA, I. (1962). An experimental investigation of the effects of attitudinal factors on eyeblink conditioning. Unpublished Ph.D. thesis, University of London.
- BRAUN, H. W. & GEISELHART, R. (1959). Age differences in the acquisition and extinction of the conditioned eyelid responses. *J. exp. Psychol.* **57**, 386-388.
- EYSENCK, H. J. (1955). Dynamic theory of anxiety and hysteria. *J. ment. Sci.* **101**, 28-51.
- EYSENCK, H. J. (1957). *The Dynamics of Anxiety and Hysteria*. London: Routledge & Kegan Paul.
- EYSENCK, H. J. (1959). *Manual of the Maudsley Personality Inventory*. London: University of London Press.
- EYSENCK, H. J. (1962). *Conditioning and personality*. *Brit. J. Psychol.* **53**, 299-305.
- EYSENCK, H. J. (1963). *Experiments in Motivation*. Oxford: (in press).
- FIELD, J. G. & BRENGELMANN, J. C. (1961). Eyelid conditioning and three personality parameters. *J. abnorm. soc. Psychol.* **63**, 517-523.
- FRANKS, C. M. (1954). An experimental study of conditioning as related to mental abnormality. Unpublished Ph.D. thesis, University of London.
- FRANKS, C. M. (1957). Personality factors and rate of conditioning. *Brit. J. Psychol.* **48**, 119-126.
- FRANKS, C. M. (1956). Conditioning and personality: a study of normal and neurotic patients. *J. abnorm. soc. Psychol.* **52**, 143-150.
- HULL, C. L. (1943). *Principles of Behaviour*. New York: D. Appleton-Century.
- JENSEN, A. R. (1958). The Maudsley Personality Inventory. *Acta Psychol.* **14**, 314-325.
- KLINGER, B. I. & PROKASY, W. F. (1962). A note on MAS score and the ready signal in classical eyelid conditioning. *Psychol. Rep.* **10**, 829-830.
- KING, M. S., KIMBLE, G. A., GORMAN, J. & KING, R. A. (1961). Replication report: two failures to reproduce effects of anxiety on eyelid conditioning. *J. exp. Psychol.* **62**, 532-533.
- MARTIN, B. (1961). The assessment of anxiety by physiological and behavioural measures. *Psychol. Bull.* **58**, 234-255.
- O'CONNOR, N. & RAWNSLEY, K. (1959). Two types of conditioning in psychotics and normals. *J. abnorm. soc. Psychol.* **58**, 157-161.
- SPENCE, K. W. & FARBER, I. E. (1953). Conditioning and extinction as a function of anxiety. *J. exp. Psychol.* **45**, 116-119.



- SPENCE, K. W. & ROSS, L. E. (1959). A methodological study of the form and latency of eyeblink responses in conditioning. *J. exp. Psychol.* **58**, 376-381.
- SPENCE, K. W. & TAYLOR, J. (1951). Anxiety and strength of the UCS as determiners of the amount of eyelid conditioning. *J. exp. Psychol.* **42**, 183-188.
- SPENCE, K. W. & TAYLOR, J. (1953). The relation of conditioned response strength to anxiety in normal, neurotic and psychotic subjects. *J. exp. Psychol.* **45**, 265-272.
- TAYLOR, J. (1951). The relationship of anxiety to the conditioned eyelid response. *J. exp. Psychol.* **41**, 81-92.
- WILLETT, R. A. (1960). Measures of learning and conditioning. In *Experiments in Personality*, vol. I (ed. Eysenck, H. J.). London: Routledge & Kegan Paul.

*Manuscript received 17 December 1962*

## Intelligence and Childhood Psychiatric Disorder

By MICHAEL RUTTER

*M.R.C. Social Psychiatry Research Unit, Maudsley Hospital,  
London, S.E.5*

The literature on the relation between intelligence and psychiatric disorders of childhood is reviewed, and an analysis is given of the distribution of I.Q.s of children first attending the Maudsley Hospital Children's Department in 1955. The distribution of I.Q.s differed significantly from that of the general population, but the differences are believed to be artifacts of referral policy. Verbal performance discrepancies were no larger than those found in the general population. Abnormality of intelligence does not seem to be a factor of major importance in child psychiatric disorders, but is worthy of attention for other reasons.

Although the relationship between the level of intelligence and delinquency has been much studied (Woodward, 1955*a, b*), the relationship of intelligence to non-delinquent psychiatric disorders of childhood has been less thoroughly explored. Nevertheless, firmly held clinical impressions regarding the relationship have been expressed and a number of studies purporting to demonstrate the role of intelligence in childhood behavioural disorders have been published. Some have found an increased incidence of disorders at both ends of the distribution of intelligence (Haggerty, 1925; McClure, 1929; Hewitt & Jenkins, 1946; Burt & Howard, 1952). Others have found relationships only with low intelligence (Blatz & Bott, 1927; Vernon, 1937; Macfarlane *et al.*, 1954), or have emphasized the role of high intelligence (Regensburg, 1926).

The relationships found have sometimes applied only to boys (Blatz & Bott, 1927) or have varied according to the ages of the children (McClure, 1929; Ackerson, 1931, 1942; Vernon, 1937; Holman, 1953). Personality and emotional problems have been noted to be associated with high intelligence, and delinquency with low (Levy, 1931; Ackerson, 1931, 1942; Hersov, 1960). Yet Hewitt & Jenkins (1946) found an association between over-inhibited behaviour and low I.Q., and unsocialized behaviour and high I.Q. Different cognitive functions have rarely been considered, but in one study emotionally disturbed children did particularly poorly on visual-motor tests (Granick, 1955). Academic achievement is often low even when the level of intelligence is normal (Salzinger, 1957). Davids (1958), on the basis of finding a low correlation between the intelligence of neurotic and schizophrenic children and their mothers, concluded that severe emotional disturbance has a pronounced detrimental effect on the intellectual functioning of children.

Thus findings are contradictory and it may be seen that the investigations on which they were based were frequently lacking in rigour. Most of the studies were carried out over 30 years ago, using tests lacking in standardization and subsequently abandoned (e.g. Haggerty, 1925; Blatz & Bott, 1927; McClure, 1929). The tests used were not always specified and results from several tests of unestablished comparability and having different standard deviations were sometimes pooled.



Measures of 'undesirable behaviour' (Haggerty, 1925) or 'misdemeanors' (Blatz & Bott, 1927) in the classroom (including items such as 'uncleanliness') may have little in common with symptom-measures concerning children attending psychiatric clinics (e.g. Levy, 1931; Ackerson, 1931, 1942; Hewitt & Jenkins, 1946). Children in the lower range of intelligence were sometimes excluded (Ackerson, 1931, 1942), only children at the extremes of the range compared (Burt & Howard, 1952) or comparisons made between crude groups, such as those above or below I.Q. 100 (Paynter & Blanchard, 1929). Artifacts due to referral biases have not received attention and in most studies tests of statistical significance were not applied.

There has been little concern regarding abnormalities of intelligence other than level of I.Q. However, Maxwell (1961), in a well-planned study, compared the cognitive test scores on the WISC battery for a sample of children attending the Maudsley Hospital with those for a sample of normal children. The standard deviations, correlations and factor loadings for the psychiatric group were considerably higher and he put forward a genetic hypothesis to explain the differences (Maxwell, 1960).

In view of the claimed association between level of intelligence and childhood psychiatric disorder and the unsatisfactory evidence upon which such claims are based, further study was thought necessary. The hypothesis tested was that a relative preponderance of the extremes of the distribution of intelligence would be found among psychiatrically ill children. Also, in so far as abnormality of intelligence is a factor in the genesis of psychiatric disorder, abnormality might be found not only in the level of general intelligence but also in the discrepancies between the levels of different abilities. A crude measure of this is the verbal-performance discrepancy in the scores on the Wechsler Intelligence Scale for Children (WISC).

## METHOD

Study was made of all new cases seen in the Maudsley Hospital Children's Department during the year 1955, this year being chosen as the most recent year in which nearly all cases had been closed at the time the analysis was undertaken. In this clinic psychometric testing was done routinely, although some children seen as emergencies and a few others, for various reasons, were not tested. The WISC was generally given but the Revised Stanford-Binet scale and Merrill-Palmer tests were used with younger children. The scores of the last, and also the Matrices which were used in two instances, were converted to a scale having a mean of 100 and standard deviation of 15, in order for them to be comparable. However, the Stanford-Binet is not directly comparable with the WISC and Wechsler-Bellevue, even when allowance is made for the difference in standard deviation (Weider *et al.*, 1943, 1951). Accordingly, the conversion table given by Weider *et al.* based on the formula  $y = 0.85X + 11$ , where  $y$  = the WISC score and  $X$  = the Binet, was used to convert the Stanford-Binet scores to the WISC equivalent. In a few children other tests such as the Goodenough Draw-A-Man Test or the Leiter scale were used, but as these are less reliable and show less good correlation with the WISC or Binet (Anastasi, 1954), such cases were not included. The tests used in the group studied are shown in Table 1.

Symptoms of the child's disorder were recorded under fields of disturbance according to a scheme based on Cameron's classification (1955). Where symptoms were predominantly those of socially disapproved conduct (including delinquency), the disorder was termed 'conduct behaviour disorder' (CBD). Those cases where symptoms were mainly in the fields of primary or secondary habit disorder, motor disorder, somatic disorder of functional origin, psychic disorder, or allergic disorder were diagnosed as 'neurotic behaviour disorder' (NBD). Symptoms of educational or work disturbance or disturbed interpersonal relationships might be included as either 'neurotic' or 'conduct' according to context. Classification

Table 1. *Intelligence tests*

Test	Number
WISC	248
WISC (one scale only)	11
Stanford-Binet	36
Merrill-Palmer	9
Matrices	2
Total	306
Not tested or test unsuitable	68

was based on overt symptoms, using the whole clinical state, not just the presenting complaints. Disorders where both neurotic and conduct disturbances were prominent, with neither predominant, were termed 'mixed behaviour disorder' (MBD). Syndromes approximating to those found in adults (anxiety state, obsessional neurosis, hysteria, depressive illness, etc.) were classified separately as 'neurotic illness' (NI). The method of symptom classification and diagnosis has been more fully described elsewhere (Rutter, 1963).

Study was confined to children in these diagnostic groupings. Children in whom such disorders were complicated by epilepsy but not directly caused by it (that is the disorder was not a manifestation of any part of an epileptic attack) were included. This was on the basis of Grunberg and Pond's finding (1957) that the background (genetic and environmental) of epileptic children with conduct disorders resembled that of a non-epileptic group of children and was strikingly different from that of a group of epileptic children without conduct disorder. The twenty-two children with uncomplicated epilepsy were excluded, as were seventeen with organic disorders and twelve in whom no psychiatric abnormality was found. Cases of mental defect and educational backwardness were included, with the exception of twenty-nine in whom this was unaccompanied by any neurotic or behaviour disorder.

## RESULTS

The distribution of intelligence among the 306 children is apparently bimodal, but comparison with the distribution expected in a standard population, the percentages being those given in the WISC manual (Wechsler, 1949), shows that the chief abnormality was an excess in the 80-89 I.Q. group with deficit in the 90-99 group (Table 2), the difference being significant at the 0.001 level. There was a smaller deficiency in the 110-119 group and an excess at both the lower and upper ends of the distribution. The distribution was similar, for those tested on the WISC and for those tested on other scales.

Table 2. *Distribution of intelligence*

I.Q.	Maudsley children		Standard population Per cent
	No.	Per cent	
Less than 69	13	4.3	2.2
70-79	26	8.5	6.7
80-89	75	24.5	16.1
90-99	54	17.7	25.0
100-109	70	22.9	25.0
110-119	37	12.0	16.1
120-129	22	7.2	6.7
More than 130	9	2.9	2.2
Total	306	100.0	100.0

$$(\chi^2 = 32.036, 7 \text{ d.f., } p < 0.001)$$



There was no significant association between the child's diagnosis and the distribution of intelligence (Table 3), nor did those with symptoms of school or work disturbance have a distribution of intelligence which differed from the group as a whole. However, there was an association with the agency of referral. A significant excess of children with an I.Q. of 79 or less occurred among those referred by the L.C.C. Children's Care Committee or a Local Education Authority ( $\chi^2 = 30.86$ , 5 d.f.,  $p < 0.001$ ). This might be expected as the children had generally been referred because of complaints by the school of the child's behaviour or progress, although it has already been noted that an excess of those of low I.Q. beyond that of the group as a whole was not found in those having symptoms of school disturbance, if other referral agencies were also included. The deficiency of I.Q. 90-99 was largely confined to those referred by the Courts or Probation Service, this group having also an excess of children with an I.Q. of 80-89 ( $\chi^2 = 21.19$ , 5 d.f.,  $p < 0.001$ ).

Table 3. *Distribution of intelligence according to diagnosis*

I.Q.	Diagnosis* (per cent)			
	NBD	NI	CBD	MBD
Less than 69	4.4	2.7	4.7	4.1
70-79	5.3	8.1	10.2	12.2
80-89	28.3	10.9	28.1	18.4
90-99	17.7	19.0	15.0	22.5
100-109	18.5	24.5	25.3	26.5
110-119	13.2	24.5	7.4	10.2
120-129	11.5	5.4	5.6	2.0
More than 130	0.9	5.4	3.7	4.1
Total No.	113	37	107	49

$$(\chi^2 = 25.116, 21 \text{ d.f.}, p < 0.3)$$

\* Key: NBD = Neurotic Behaviour Disorder; NI = Neurotic Illness; CBD = Conduct Behaviour Disorder; MBD = Mixed Behaviour Disorder.

No significant differences were found in the distribution of intelligence between the different age groups and there was no significant variation according to the sex of the child. The excess of I.Q.s 80-89 was greatest for the children aged 7 years or less, but the difference from other age groups was small and the chi-square did not reach the 5 per cent level of significance. There was a significant and positive relationship between social class (according to the Registrar General's classification) and intelligence ( $\chi^2 = 34.29$ , 15 d.f.,  $p < 0.01$ ). The chief contributions to the chi-square were the excess of children with an I.Q. of 120 or more in classes I and II (professional and business) and the corresponding deficit of those scoring in the 80-89 range. As this is in keeping with findings in the general population (Anastasi, 1954), there is no support for the view that psychiatric disorder is particularly common in the dull child from the upper-class home and the bright child from the lower-class home.

The severity of the disorder was roughly estimated according to the number of symptoms (children with neurotic illness of adult type being excluded) and the extent by the number of fields of disturbance involved. There was no significant variation of severity according to the level of intelligence (Table 4), nor was there any relationship between I.Q. and the extent of the disorder.

A significant excess ( $p < 0.01$ ) of those with an I.Q. of 79 or less were not accepted for treatment after being seen once as an out-patient (14 out of 39 compared with 45 out of 267 among the remainder of the children), but of those taken on for treatment the duration of out-patient

Table 4. *Distribution, according to intelligence, of number of symptoms per patient*

Number of symptoms	I.Q. (per cent)					
	Less than 79	80-89	90-99	100-109	110-119	More than 120
1-5	37.1	39.5	36.2	41.0	42.9	29.6
6-9	37.1	35.2	31.9	41.0	46.4	59.3
More than 10	25.8	25.3	31.9	18.0	10.7	11.1
Total No.	35	71	47	61	28	27

$(\chi^2 = 11.856, 10 \text{ d.f.}, p < 0.5)$

care was not related to I.Q. nor did the number of psychiatrist-patient interviews vary with the intelligence level. The outcome of treatment was routinely recorded in the case records by the psychiatrist treating the child. Such an assessment has little meaning unless further specified in terms of alteration of symptoms or social adjustment. Further, any change that may have taken place cannot be attributed to treatment without further study. Nevertheless, although without value for comparison with other clinics, such evaluations may be of limited value when used to compare groups of cases within the same clinic, and it is in this way that they are used here. Comparison of outcome in the different I.Q. groups revealed no significant differences (Table 5) and so the data did not suggest that the level of improvement was related to the child's I.Q.

Table 5. *Distribution of outcome according to intelligence*

Outcome	I.Q. (Per cent)					
	Less than 79	80-89	90-99	100-109	110-119	More than 120
Worse or no change	28.0	13.0	8.7	26.7	39.4	17.2
Improved or slightly improved	48.0	51.8	41.3	43.4	27.3	34.5
Much improved or recovered	20.0	25.9	43.5	23.3	30.3	41.4
Not closed or not known	4.0	9.3	6.5	6.6	3.0	6.9
Total No.	25	54	46	60	33	29

$(\chi^2 = 22.26, 15 \text{ d.f.}, p < 0.2)$

#### *Verbal-performance discrepancies (on the WISC)*

Cases of children who had been tested on the WISC were further examined for discrepancies between the scores on the verbal and performance scales. In the group as a whole the proportion where the verbal score exceeded the performance score was similar to that where the converse occurred (Table 6) and there was no significant variation according to the child's diagnosis.

Field (1960) has provided tables whereby the significance of any given verbal-performance discrepancy may be calculated according to its frequency in the general population. From this table giving the proportion in a standard population having different discrepancies (Table 7), the expected proportions of any population of known age on the different groupings according to the degree of discrepancy can be calculated. This was done for the group studied, taking into account the age of each patient, and converting Field's cumulative percentages into percentile groupings which summated to 100 per cent (Table 8). There was a slight tendency for the observed larger discrepancies to occur more frequently than expected, and vice-versa, but the differences fell short of significance. There was a tendency for the differences to be



Table 6. *Distribution according to diagnosis of verbal-performance scale ratio*

Scores	NBD (per cent)	NI (per cent)	CBD (per cent)	MBD (per cent)	All children (per cent)
Verbal greater than performance	47.0	60.6	49.4	43.6	49.2
Equal scores	6.0	3.0	5.4	00.0	4.4
Performance greater than verbal	47.0	36.4	45.2	56.4	46.4
Total No.	83	33	93	39	248

( $\chi^2$  not significant)

more marked in those with school disturbance than in those without, but the differences were not significant ( $p < 0.2$ ). Similarly the differences were most marked in the 10-13 year age group, but when compared with the rest, the differences again fell short of significance ( $p < 0.2$ ). However, there was a significant variation according to diagnosis, in that there was an increased frequency of the larger verbal-performance discrepancies among those diagnosed as having a neurotic illness or a mixed behaviour disorder.

Table 7 (from Field, 1960). *The Abnormality of verbal-performance discrepancies*

Per cent in population obtaining a given or greater discrepancy	Age (years)		
	7½	10½	13½
50	9.0	8.0	9.4
25	15.4	13.8	16.2
20	17.2	15.4	18.0
10	22.0	19.7	23.1
5	26.3	23.5	27.6
2	31.2	28.0	32.8
1	34.6	31.0	36.3
0.1	44.1	39.5	46.3

The figures in the body of the table refer to verbal-performance discrepancies in either direction. Thus, among children aged 10½ years 25 per cent have discrepancies of 13.8 or greater, 10 per cent. discrepancies of 19.7 or greater, etc.

Table 8. *Distribution of verbal-performance discrepancies by percentile groupings*

Observed		Expected	
No.	Per cent	No.	Per cent
111	44.8	124.0	50.0
66	28.6	62.0	25.0
43	15.3	37.2	15.0
10	4.1	12.4	5.0
13	5.2	7.4	3.0
5	2.0	5.0	2.0

( $\chi^2 = 7.23$ , 5 d.f.,  $p < 0.3$ )

## DISCUSSION

Among children first attending the Maudsley Hospital Children's Department in 1955 the distribution of intelligence differed significantly from normal. There was a slight excess at both ends of the distribution, but the most striking abnormality was the excess of children with an I.Q. in the range 80-89, there being a corresponding deficit in the range 90-99.

As part of the Department there is a clinic for mentally subnormal children. Although such children were excluded if they exhibited no abnormality beyond that of low intelligence (being included, however, if other symptoms were present), an increased association between the mental subnormality and psychiatric disorder in the clinic population would be expected on purely statistical grounds. Berkson (1946) has clearly shown that, for reasons separate from referral biases, the ratio of multiple diagnoses to single diagnoses in hospitals will always be greater than in the general population, this artifact being found in the association between conditions where only a small proportion of each are referred to hospital—a condition applying to both mental subnormality and childhood psychiatric disorder. This might explain the excess of children in the group with the lowest levels of intelligence, and it seems fair to conclude that severe mental subnormality is not of great importance in the aetiology of childhood behavioural and neurotic disorders. However, this may well not be so in areas of the world where ascertainment of and schooling for the intellectually handicapped is less comprehensive (Priestley, 1959). Nor, indeed, may it have been true in this country in earlier years. It should also be stated that, although apparently not an important factor in causing disorders of personality or behaviour, mental subnormality is a condition associated with other defects of functioning and, in its own right, merits psychiatric attention.

The distribution of intelligence showed no significant variation with the child's diagnosis, but the deficiency of I.Q.s in the range 90-99 was largely accounted for by those children referred from the Courts—that is the delinquent group, but not the group with anti-social symptoms who had not been before the Courts. A referral bias seems likely and, as the deficiency of I.Q.s 90-99 is largely restricted to children referred by the Courts, not much weight can be attached to this finding for the group as a whole at the moment.

The excess at both ends of the distribution of intelligence was slight, but in so far as it was of any importance, it suggested that psychiatric disorder was associated with intellectual abilities differing from normal rather than with mental subnormality as such. This is in keeping with most previous studies but, in contrast, the present investigation provides no evidence that the level of intelligence was related to the form of psychiatric disorder exhibited by the child.

Verbal-performance discrepancies in those children tested on the WISC (this being the majority) were not very remarkable. The slightly (but not significantly) increased size of discrepancies would also be in keeping with the view that differences from the normal, in any direction, may be more likely to be associated with disorder in other spheres. There was no evidence that the direction of verbal-performance discrepancy was of any importance. Of course, it might well be so in the individual



case, but overall neither direction of discrepancy was more common. The relationship with diagnosis was unexpected and, although statistically significant, the psychological significance of the finding is not clear.

Overall, there was no significant abnormality of verbal-performance discrepancy either in size or direction, and it may be concluded that such abnormalities showed little association with childhood psychiatric disorder. Further, the level of intelligence in the clinic population, although differing significantly from normal, differed to only a relatively slight extent, this being probably not a factor of major aetiological importance. The findings refer to one hospital only but other studies also have generally found only slight associations between intelligence and psychiatric disorder. Although mainly negative, results have been presented in some detail for methodologic reasons. In view of the positive findings in some earlier studies, the apparent bimodality in the overall distribution of I.Q. in the present study could have been interpreted as support for such findings. Reasons have been given for concluding that the relationship is largely artifactual.

Children with an I.Q. of 79 or less in this clinic were less likely to be taken on for treatment, but otherwise intelligence played no part in their selection. There was no evidence that such children showed a lesser degree of disturbance—indeed there was no significant relationship between the level of intelligence and the severity or extent of the disorder, as measured by the number of symptoms and fields of disturbance respectively. Further, there did not appear to be any significant relationship between the intelligence of the child patient and either the course of treatment or its outcome.

In conclusion, abnormalities of intelligence were found not to be an important factor in either the genesis or course of treatment of behavioural or neurotic disorder in childhood. Children of differing intellectual abilities may have to face somewhat different issues in development and it is of aid in the assessment and treatment of children to have knowledge of their capabilities and attainments. Although the present findings suggest that one should be very cautious in attributing behavioural or personality disorders to abnormalities of intelligence *per se*, nevertheless children having such abnormalities are an important group, well deserving of psychiatric notice through the presence of other defects of functioning and their concomitants.

### SUMMARY

1. The literature on the relationship between intelligence and non-delinquent psychiatric disorders of childhood was reviewed.
2. A study of the intelligence of children first attending the Maudsley Hospital in 1955 was described.
3. The distribution of intelligence differed significantly from normal, there being an excess of I.Q.s 80-89, a deficit within the ranges 90-99 and 110-119, and lesser excesses at both ends of the distribution.
4. The deficiency of I.Q.s 90-99 was largely confined to those referred by the Courts and the excess of those of low intelligence was greatest among those referred by the L.C.C. Children's Care Committee or a Local Education Authority.
5. Reasons were given for supposing that the excess of those of low intelligence could be largely explained as due to an artifact.

6. Verbal-performance discrepancies on those children tested on the WISC did not differ significantly in either size or direction from those found in the general population.

7. Those of lowest intelligence were less likely to be taken on for treatment, but otherwise intelligence was not related to the duration, frequency or outcome of treatment, nor to the likelihood of in-patient care.

8. It was concluded that abnormality of intelligence was not a major factor of importance in the aetiology of childhood behavioural or neurotic disorders, but that it was worthy of psychiatric attention for other reasons.

I am grateful to Dr Kenneth Cameron for his advice and for permission to use the facilities of the Maudsley Hospital Children's Department. I am also indebted to Dr A. E. Maxwell for assistance with statistical matters, and to Dr H. G. Birch for constructive and helpful comments on an earlier draft.

### REFERENCES

- ACKERSON, L. (1931). *Children's Behaviour Problems, I*. Chicago: University Press.
- ACKERSON, L. (1942). *Children's Behaviour Problems, II*. Chicago: University Press.
- ANASTASI, A. (1954). *Psychological Testing*. New York: Macmillan.
- BERKSON, J. (1946). Limitations of the application of fourfold table analysis to hospital data. *Biometrics*, **2**, 47-59.
- BLATZ, W. E. & BOTT, E. A. (1927). Studies in mental hygiene of children I. Behaviour of public school children—a description of method. *Ped. Sem.* **34**, 552-582.
- BURT, C. & HOWARD, M. (1952). The nature and causes of maladjustment among children of school age. *Brit. J. Psychol. statist. Sect.* **5**, 39-59.
- CAMERON, K. (1955). Diagnostic categories in child psychiatry. *Brit. J. med. Psychol.* **28**, 67-71.
- DAVIDS, A. (1958). Intelligence in childhood schizophrenics, other emotionally disturbed children, and their mothers. *J. cons. Psychol.* **22**, 159-163.
- FIELD, J. C. (1960). Two types of tables for use with Wechsler's intelligence scales. *J. clin. Psychol.* **16**, 3-7.
- GRANICK, S. (1955). Intellectual performance as related to emotional instability in children. *J. abnorm. (soc.) Psychol.* **51**, 653-656.
- GRUNBERG, F. & POND, D. A. (1957). Conduct disorders in epileptic children. *J. Neurol. Psychiat.* **20**, 65-68.
- HAGGERTY, M. E. (1925). The incidence of undesirable behaviour in public school children. *J. educ. Res.* **12**, 102-111.
- HERSOV, L. A. (1960). Persistent non-attendance at school. *J. Child Psychol. Psychiat.* **1**, 130-136.
- HEWITT, L. E. & JENKINS, R. L. (1946). *Fundamental Patterns of Maladjustment—The Dynamics of their Origin*. State of Illinois.
- HOLMAN, P. (1953). Some factors in the aetiology of maladjustment in children. *J. ment. Sci.* **99**, 654-688.
- LEVY, J. (1931). A quantitative study of the relationship between intelligence and economic status as factors in the aetiology of children's behaviour problems. *Amer. J. Orthopsychiat.* **1**, 152-162.
- MACFARLANE, J. W., ALLEN, L. & HONZIK, M. P. (1954). *A Developmental Study of the Behaviour Problems of Normal Children between 21 months and 14 years*. Berkeley and Los Angeles: University of California Press.
- MAXWELL, A. E. (1960). Discrepancies in the variances of test results for normal and neurotic children. *Brit. J. Psychol. statist. Sect.* **13**, 165-172.
- MAXWELL, A. E. (1961). Discrepancies between the patterns of abilities for normal and neurotic children. *J. ment. Sci.* **107**, 300-307.
- MCCLURE, W. E. (1929). Characteristics of problem children based on judgments of teachers. *J. juv. Res.* **13**, 124-140.



- MUNDY, L. & MAXWELL, A. E. (1958). Assessment of the feeble-minded. *Brit. J. med. Psychol.* **21**, 201-210.
- PAYNTER, R. H. & BLANCHARD, P. (1929). *Educational Achievement of Problem Children*. New York: The Commonwealth Fund.
- PRIESTLEY, K. E. (1959). *Constructive Work for Mental Health in Hong Kong* (ed.). London: Oxford University Press.
- REGENSBURG, J. (1926). Emotional handicaps to intellectual achievement in supernormal children. *Ment. Hyg., Lond.*, **10**, 480-494.
- RUTTER, M. L. (1963). Illness in parents and children. M.D. Thesis, University of Birmingham. (Version for publication in preparation.)
- SALZINGER, K. (1957). Academic achievement in a group of mentally disturbed adolescents in a residential treatment setting. *J. genet. Psychol.* **90**, 239-253.
- VERNON, P. E. (1937). A study of the norms of the validity of certain mental tests at a child guidance clinic. *Brit. J. educ. Psychol.*, **7**, 72-88; 115-137.
- WECHSLER, D. (1949). *Wechsler Intelligence Scale for Children*. New York: The Psychological Corporation.
- WEIDER, A., LEVI, J. & RISCH, F. (1943). Performances of problem children on the Wechsler-Bellevue intelligence scales and the revised Stanford-Binet. *Psychiat. Quart.* **17**, 696-701.
- WEIDER, A., MOLLER, P. A. & SCHRAMM, T. A. (1951). The Wechsler Intelligence Scale for Children and the Revised Stanford-Binet. *J. cons. Psychol.* **15**, 330-333.
- WOODWARD, M. (1955). *Low Intelligence and Delinquency*. London: I.S.T.D. Publ.
- WOODWARD, M. (1955). The role of low intelligence in delinquency. *Brit. J. Delinq.* **5**, 280-294.

*Manuscript received 22 February 1963*

## The Reliability of the Pressure Algometer

BY H. MERSKEY AND F. G. SPEAR

*University Department of Psychiatry  
Whiteley Wood Clinic, Sheffield*

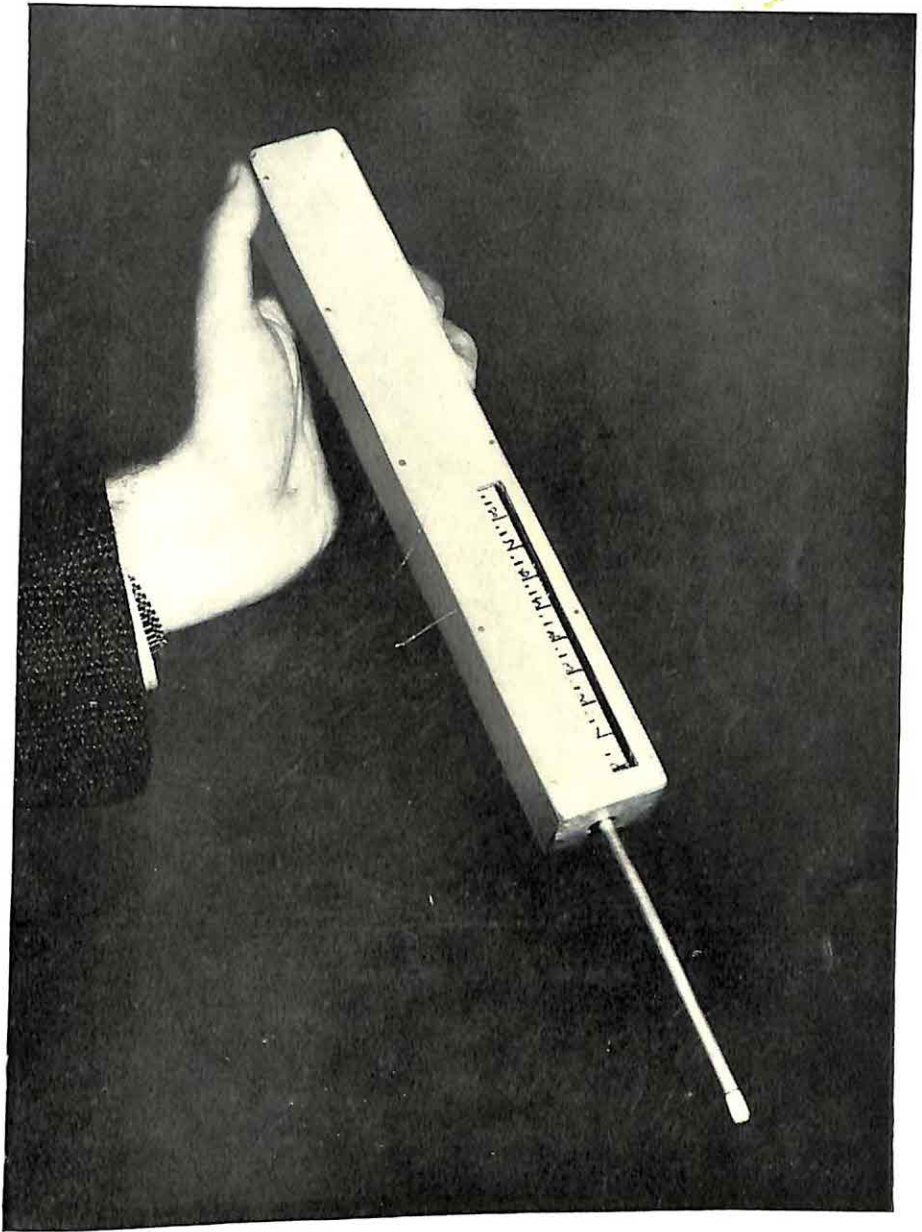
The pressure algometer is a simple clinical instrument which has been used to measure the response to pressure pain in normal subjects. The point at which subjects first said they felt pain was called the Verbal Report of Pain (VRP). The point at which they said the pressure 'hurt a lot' was called the Pain Reaction Point (PRP). Both points were significantly lower in women than in men, but there was no significant difference between white male medical students, Afro-Asian male medical students and white male student teachers. The method was shown to have a moderate degree of consistency and reliability when used independently by two observers. The VRP and PRP were highly correlated with each other, suggesting that both were a measure of the same tendency.

### INTRODUCTION

A reliable and simple method of measuring the response to pain would be of much use, both in clinical practice and experimentally. We report here an investigation of the reliability of one of the oldest of such methods—the use of the pressure algometer. This consists of a plunger mounted on a calibrated spring (Plate 1). The plunger has a flat end which can be applied to suitably shaped bony surfaces such as the forehead and tibia. It has been used since Victorian times (Keele, 1954). Head & Holmes (1911) called it 'Cattell's Algometer' and employed it to study the thalamic syndrome. Related methods have been used by Libman (1934), who assessed individual sensitivity from the response to styloid pressure, by Hollander (1939), who used a rough metal grater under an inflatable cuff, and by Perner (1941), who modified a Geneva Lens Measure. This provided another system of direct recording of mechanical pressure applied to a bony surface (actually the proximal phalanx of the thumb). Gluzek (1944) similarly introduced a quantitative method of measuring pain and deep sensibility. He used discs forced against the skin by air pressure. Lastly, Clutton-Brock (1957) has employed pressure from a round-headed screw attached to a spring balance.

The method of Hollander was tried out by Wilder (1940) and that of Libman by Perner who compared it with his own technique; Sherman (1943) compared Libman's and Hollander's methods, and Keele compared both of these with each other and with the pressure algometer. From these studies generally equivalent results emerge for the different methods of estimating sensitivity and Wilder and Sherman, who each examined males and females, neurotics and others, agree in finding that women are more sensitive than men and neurotics more sensitive than patients with 'organic' complaints. These findings are further in harmony with other studies of pain sensitivity (Chapman *et al.*, 1947; Schilling & Musser, 1949; Hemphill *et al.*, 1952; Kennard, 1952; Hall & Stride, 1954; Dundee & Moore, 1960). Keele suggested that the pressure algometer was preferable to Libman's method since it gave





A pressure algometer.





directly calibrated results. In this it corresponds to Pelner's sensometer, Gluzek's dolorimeter and Clutton-Brock's analgesiometer. None of these authors has provided direct evidence of the reliability of his technique, although each one mentions or suggests that reliability is an advantage of the method described. A range of variation for individual readings with Clutton-Brock's technique is specified, however, by Dundee and Moore, and the reliability of the pressure algometer between different observers in certain circumstances was assessed by Merskey *et al.*, (1962), using schizophrenic subjects. We report here a study of the reliability of the pressure algometer in normal subjects, using different observers and stimulus sites.

### DESIGN OF INVESTIGATION

An algometer was used of the pattern described by Head & Holmes (1911), Keele and Merskey *et al.*, having a flat circular end of 0.5 cm diameter and calibrated from 0 to 7.75 kg. A total of sixty-nine normal subjects were examined independently by two observers (H.M. & F.G.S.). Forty-nine of these subjects were an unselected series of medical students examined during their appointment as psychiatric clerks. Because sex and racial or cultural differences are sometimes held to influence the threshold or response to pain, these subjects have been considered in three groups, viz.: Group 1, twenty-eight white male medical students; Group 2, eleven Afro-Asian male medical students; and Group 3, ten white female medical students. The remaining twenty subjects, constituting Group 4, were twenty white male post-graduate students studying for the Diploma in Education. These were seen during routine medical examination. All subjects were told that the procedure was for research purposes and only one of those approached refused to take part. With one exception all the subjects were under 30 years of age, the mean being 22.8 years. Of the forty-nine subjects in Groups 1-3, twenty-six were examined first by Observer 1 and twenty-three were examined first by Observer 2. All the subjects in groups 1-3 were examined on the forehead twice. The twenty subjects in Group 4 were also examined twice but on one occasion over the tibia and on another occasion on the forehead. Equal numbers in this last group were examined first on the shin and first on the forehead, and by each observer. Thus Observer 1 examined five subjects first on the shin and five subjects on the forehead, and Observer 2 did likewise.

### Technique

All the subjects were told 'This is a test of sensitivity, not a test of endurance'. In the first three groups they were then asked to say (a) when pressure from the instrument began to hurt, and (b) when it hurt a lot. After a trial run four pairs of readings were taken on each occasion, giving four estimates of the 'Verbal Report of Pain' (VRP)—when pain began, and the 'Pain Reaction Point' (PRP)—when it hurt a lot.

In the fourth group the subjects were only asked to say 'When it hurt a lot'. In all cases the attempt was made to increase pressure smoothly at a consistent rate of about 1 kg/sec as recommended by Keele.

### RESULTS

Except in Table 6 all the results have been computed from the averages of the four readings taken from each subject on each occasion for the 'Verbal Report of Pain' (where measured) and the 'Pain Reaction Point'. Table 1 shows mean values for the VRP. These do not differ significantly between occasions or observers although there is a tendency for Observer 1 to score higher than Observer 2 and for subjects to score higher on Occasion 1 than on Occasion 2. The female students, however, scored significantly lower than either group of males. Similar trends and similar significant differences for female students appear with the measurement of the Pain Reaction Point set out in Table 2. The finding of the sex differences is in accordance

Table 1. *Verbal report of pain for different groups observers and occasions: means and standard deviations and group coefficients of variation*

Groups	N	Observer		Occasion		Group averages	Coefficient of Variation (S.D./Mean $\times 100$ )
		1	2	1	2		
1. White male medical students	28	3.95 $\pm$ 1.06	3.54 $\pm$ 0.81	3.92 $\pm$ 1.10	3.57 $\pm$ 0.77	3.75 $\pm$ 0.82	22%
2. Afro-Asian male medical students	11	4.01 $\pm$ 0.87	3.18 $\pm$ 1.38	4.15 $\pm$ 1.34	3.53 $\pm$ 0.84	3.84 $\pm$ 1.01	26%
3. White female medical students	10	2.86 $\pm$ 0.85	2.63 $\pm$ 1.21	2.99 $\pm$ 1.19	2.50 $\pm$ 0.81	2.74 $\pm$ 0.91	33%
Groups 1-3	49	3.74 $\pm$ 1.06	3.38 $\pm$ 1.09	3.78 $\pm$ 1.22	3.34 $\pm$ 0.89		

Units: Kg/unit area of 0.1963 cm<sup>2</sup>.\* *t*-test:  $p < 0.05$  for difference between Group 2 and Group 3.† *t*-test:  $p < 0.01$  for difference between Group 1 and Group 3.Table 2. *Pain reaction points for different groups, observers and occasions: means and standard deviations and group coefficients of variation*

Groups	N	Observer		Occasion		Group averages	Coefficient of variation (S.D./Mean $\times 100$ )
		1	2	1	2		
1. White male medical students	28	6.25 $\pm$ 1.26	5.76 $\pm$ 1.25	6.19 $\pm$ 1.29	5.82 $\pm$ 1.25	6.01 $\pm$ 1.17	19%
2. Afro-Asian male medical students	11	6.05 $\pm$ 1.2	5.6 $\pm$ 1.58	5.94 $\pm$ 1.43	5.71 $\pm$ 1.40	5.83 $\pm$ 1.33	23%
3. White female medical students	10	4.49 $\pm$ 1.59	4.28 $\pm$ 2.06	4.64 $\pm$ 1.9	4.13 $\pm$ 1.74	4.38 $\pm$ 1.74	46%
4. White male student teachers	20	5.82 $\pm$ 1.43	5.51 $\pm$ 1.59	5.7 $\pm$ 1.52	5.63 $\pm$ 1.51	5.66 $\pm$ 1.51	27%
Groups 1-4	69	5.86 $\pm$ 1.47	5.43 $\pm$ 1.55	5.79 $\pm$ 1.53	5.50 $\pm$ 1.51		

Units: Kg/unit area of 0.1963 cm<sup>2</sup>.\* *t*-test:  $p < 0.001$  for difference between Group 1 and Group 3.† *t*-test:  $p < 0.05$  for difference between Group 4 and Group 3.

with the results of Wilder (1940), Sherman (1943), Chapman & Jones (1944), Kennard (1952), Hall & Stride (1954), and Dundee & Moore (1960).

Separate figures are also available for the PRP for Group 4 with respect to forehead and shin. The mean values and standard deviations were: forehead  $5.35 \pm 1.47$ , shin  $5.98 \pm 1.49$ . These values for the forehead tend to be less than on the shin but the difference is not statistically significant, *t*-test  $0.1 < p > 0.05$ . It does seem, however, as if a larger series might show a significant difference between the pressure tolerated at the two sites.

When the figure for the VRP is subtracted from that for PRP a value is obtained which may be termed the Reaction Interval. The mean value for this is also significantly lower in the women than in the other groups (Table 3).

A measure of the reliability of the various responses by the subjects, despite



changes of observer, occasion and site, is next presented in Table 4. This Table shows that the VRP is less reliable than the PRP. The correlation figures obtained for the latter for Groups 1-3 are very similar to those obtained in another study (Merskey *et al.*, 1962), using schizophrenic subjects. In the case of Group 4 the degree of inter-correlation for PRP was lower than for the other groups even when

Table 3. *Reaction interval (PRP-VRP) for different groups*

	Medical students	N	Mean and standard deviation	Coefficient of variation (S.D./Mean $\times$ 100)
Group 1	White males	28	$\dagger 2.26 \pm 0.84$	37%
Group 2	Afro-Asian males	11	$* 1.98 \pm 0.76$	38%
Group 3	White females	10	$\dagger * 1.64 \pm 1.1$	67%

Units: Kg/unit area of  $0.1963 \text{ cm}^2$ .

\* *t*-test:  $p < 0.05$  for difference between Group 2 and Group 3.

$\dagger$  *t*-test:  $p < 0.01$  for difference between Group 1 and Group 3.

Table 4. *Product-moment correlations for the Verbal Report of Pain (VRP) and Pain Reaction Point (PRP) for different occasions, observers and sites*

	N	Correlations of VRP with VRP	Correlations of PRP with PRP
Groups 1-3			
Occasion I with Occasion II	49	$+0.65\dagger$	$+0.813\dagger$
Observer I with Observer II	49	$+0.59\dagger$	$+0.814\dagger$
Group 4			
Occasion I with Occasion II	20		$+0.63^*$
Observer I with Observer II	20		$+0.66^*$
Shin with forehead	20		$+0.71\dagger$

\* *t*-test:  $p < 0.01$ .

$\dagger$  *t*-test:  $p < 0.001$ .

Table 5. *Product-moment correlations between Verbal Report of Pain (VRP), Pain Reaction Point (PRP) and their difference (Reaction Interval—RI)*

	N	Correlations of VRP with PRP	Correlations of RI with VRP	Correlations of RI with PRP
Group 1	28	$+0.70\dagger$	$-0.002$ (N.S.)	$+0.76\dagger$
Group 2	11	$+0.82\dagger$	$+0.11$ (N.S.)	$+0.72^*$
Group 3	10	$+0.84\dagger$	$+0.50$ (N.S.)	$+0.71^*$

\* *t*-test:  $p < 0.02$ .

$\dagger$  *t*-test:  $p < 0.001$ .

these were considered separately, although still highly significant. Possible reasons for this are discussed below. From Table 5 it is evident that VRP and PRP correlate well with each other. The Reaction Interval also correlates very highly with the PRP but not with the VRP.

The possibility was considered that the first, second, third or fourth readings in each group might show a tendency to rise or fall. The ratios of the individual readings to the first one were therefore calculated for VRP and PRP. Table 6 shows that these ratios do not differ significantly from unity. There is therefore no evidence of a tendency for the response to rise or fall with successive readings taken at any one

Table 6. *Ratios of individual readings to first one in each set of four: means, standard deviations and coefficients of variation*

Ratios	VRP N = 98	Coefficients of variation	PRP N = 138	Coefficients of variation
B <sub>2</sub> /B <sub>1</sub>	1.01 ± 0.22	22%	0.99 ± 0.14	14%
B <sub>3</sub> /B <sub>1</sub>	1.02 ± 0.22	22%	0.98 ± 0.16	16%
B <sub>4</sub> /B <sub>1</sub>	0.98 ± 0.25	26%	0.95 ± 0.18	19%

1st Reading B<sub>1</sub>, 2nd Reading B<sub>2</sub>, etc.

time. Table 6 also provides a measure of the scatter of readings, given by the Standard Deviations of the ratios and their Coefficients of Variation. These last are generally lower than those in Tables 1, 2 and 3, showing less variation between different readings taken on each individual than between individuals.

### DISCUSSION

Whether there is such a thing as a measurable pain threshold is still a matter for dispute. Lanier (1943) gave evidence that a consistent threshold to electrical stimulation was nearly impossible to secure. Hardy, Wolff & Goodell (1940, 1952) claimed to find such a threshold for heat-pain, but their conclusions have been vigorously disputed by Beecher (1959). In these cases there is indeed considerable evidence mentioned by Hall (1953) that the point at which the subject first notices a pain depends greatly upon his attitude and related factors.

As with electric and heat stimulation, so with gross pressure, there is doubt whether a 'true threshold' exists. Keele (1954) assumed he was measuring a threshold and showed that the point of first response rose greatly over denervated areas but most of those investigators who have examined the response to gross pressure (Libman, 1934; Hollander, 1939; Wilder, 1940; Perner, 1941; Sherman, 1943) have concentrated on the indication it gives of the subject's emotional sensitivity rather than his sensory physiology.

In doing so these workers have also provided much evidence that the point at which the patient first notices or reports pain depends a great deal upon such factors as attitude, sex, cultural role, and so forth. We have therefore followed the practice of Hall & Stride (1954) who preferred to avoid the term 'threshold' and to speak of the Verbal Report of Pain (VRP) as the point at which the subject first describes an awareness of pain. The fact that in our study the VRP or 'threshold' correlates very highly with the PRP, which is a measure of the tolerance for pain, probably also supports the view that pain 'threshold' is determined by other factors than purely physiological ones.

It has already been noted that with this technique the VRP and PRP correlate well between different observers, occasions and sites. The scatter of the observations can be taken into account by examination of the co-efficients of variation. For the VRP these range from 22 per cent to 33 per cent. For the PRP they range from 19 per cent to 46 per cent. Such values are intermediate between those found for heat-pain (which tend to be low) and those given by Lanier (1943) for electrical stimulation



(Hall, 1953). They correspond with those found by Dundee & Moore (1960) in their studies with Clutton-Brock's method.

We think, therefore, that pressure-pain has a degree of reliability which makes possible its quantitative use for investigations of emotional states and anaesthesia and analgesia. This is particularly so where each individual serves as his own control, since the scatter of readings taken on each person tends to be less than that of readings taken on several people. It was seen that VRP and PRP are highly correlated with each other, suggesting that both are measures of the same tendency in the subject. Even though the VRP and PRP are correlated, however, it is not necessary for the interval which separates them to alter with either of these factors. As it happens, the size of the interval does vary significantly with different groups just as the VRP and PRP have varied and further the Pain Reaction Point shows a very significant correlation with the Reaction Interval.

When the correlations for PRP with itself are further examined it is noteworthy that they are higher in the medical students than in the student teachers although the increase is not significant. We think this is likely to be due either to an increase in the variance produced by incorporating different sites in the design of the investigation, or to a tendency for the PRP to be fixed more precisely at a particular point when the VRP is first sought as well. The student teachers, it will be noted, were not asked for a VRP but only for a PRP—unlike the medical student groups.

We wish, first of all, to thank our subjects for their co-operation, and Dr P. W. W. Gifford of Sheffield University Student Health Service for access to the Student-Teachers. We are grateful to Mr P. H. Blundell, Director of the Sheffield University Computing Laboratory, for programming the data and for his help in the use of the electronic computer. Additional helpful comments were made by Mr P. R. F. Clarke, Mr Jonathan Rosenhead and Professor Erwin Stengel, as well as other colleagues, to all of whom our thanks are due.

## REFERENCES

- BEECHER, H. K. (1959). *Measurement of Subjective Responses: Quantitative Effects of Drugs*. New York: Oxford University Press.
- CHAPMAN, W. P. & JONES, C. M. (1944). Variations in cutaneous and visceral pain sensitivity in normal subjects. *J. clin. Invest.* **23**, 81-91.
- CHAPMAN, W. P., COHEN, M. E. & COBB, S. (1947). Pain in neurocirculatory asthenia, combat fatigue, and anxiety neurosis. *J. clin. Invest.* **25**, 890-896.
- CLUTTON-BROCK, J. (1957). The cerebral effects of over-ventilation. *Brit. J. Anaesth.* **29**, 111-113.
- DUNDEE, J. W. & MOORE, J. (1960). Alterations in response to somatic pain associated with anaesthesia. *Brit. J. Anaesth.* **32**, 396-406.
- GLUZEK, L. J. B. (1944). Dolorimetry: A quantitative method of measuring pain and deep sensibility. *Ohio St. med. J.* **40**, 49-50.
- HALL, K. R. L. (1953). Studies of cutaneous pain. *Brit. J. Psychol.* **44**, 281-294.
- HALL, K. R. L. & STRIDE, E. (1954). The varying response to pain in psychiatric disorders: A study in abnormal psychology. *Brit. J. med. Psychol.* **27**, 48-60.
- HARDY, J. D., WOLFF, H. G. & GOODELL, H. (1940). A new method for measuring pain threshold: Observations on spatial summation of pain. *J. clin. Invest.* **19**, 649-657.
- HARDY, J. D., WOLFF, H. G. & GOODELL, H. (1952). *Pain Sensations and Reactions*. Baltimore: Williams and Wilkins.
- HEAD, H. & HOLMES, G. (1911). Sensory disturbances from cerebral lesions. *Brain*, **34**, 102-254.

- HEMPHILL, R. E., HALL, K. R. L. & CROOKES, T. G. (1952). Fatigue and pain tolerance in depressive and psychoneurotic patients. *J. ment. Sci.* **98**, 433-440.
- HOLLANDER, E. (1939). A clinical gauge of sensitivity to pain. *J. Lab. clin. Med.* **24**, 537-538.
- KEELE, K. D. (1954). The pressure algometer. *Lancet*, **1**, 636-639.
- KENNARD, M. (1952). The responses to painful stimuli of patients with severe chronic painful conditions. *J. clin. Invest.* **31**, 245-252.
- LANIER, L. H. (1943). Variability in the pain threshold. *Science*, **97**, 49-50.
- LIBMAN, E. (1934). Observations on individual sensitiveness to pain with special reference to abdominal disorders. *J. Amer. med. Ass.* **102**, 335-341.
- MERSKEY, H., GILLIS, A. & MARSZALEK, K. S. (1962). A clinical investigation of reactions to pain. *J. ment. Sci.* **108**, 347-355.
- PELNER, L. (1941). The determination of sensitivity to pain. A simple clinical method. *J. lab. clin. Med.* **27**, 248-251.
- SCHILLING, R. F. & MUSSER, M. J. (1949). Pain reaction threshold in the menopausal syndrome. *Amer. J. med. Sci.* **218**, 204-206.
- SHERMAN, E. D. (1943). Sensitivity to pain. (With an analysis of 450 cases.) *Canad. med. Ass. J.* **48**, 437-441.
- WILDER, R. M. JR. (1940). Sensitivity to pain. *Proc. Mayo Clin.* **15**, 551-554.

*Manuscript received 7 June 1963*



## A Note on the MPI: A Response Tendency in Neurotics Related to Intelligence

BY T. G. CROOKES AND S. J. HUTT\*

*St. John's Hospital, Aylesbury*

In the standardizing group Eysenck (1959) found high correlations between the short and full scales of the MPI: 0.86 for Neuroticism (N) and 0.87 for Extroversion (E). Having noted some patients who seemed to change their type of response as they progressed down the inventory, we examined the correlation in an unselected group of neurotics. Records were available of a group of 60 successive neurotic (including psychopathic) patients seen in the psychology department, collected for comparison with psychotics (Crookes & Hutt, 1963).

Since the full scale score includes the short scale, it was thought appropriate, as we wished to use tests of significance, to use the correlations between the short scales and the remainder (i.e. between the first six and the last eighteen questions on each scale), rather than the correlations between short and full scales. Eysenck gives the data from which these correlations can be calculated, and they are given in Table 1 together with those of our neurotic group. Also given are the probabilities of obtaining our correlations from a population whose correlation is that of Eysenck's group.

On both scales the correlations for neurotics are significantly lower than the normal. Inspection of the correlation charts shows that on the E-scale large deviations from the

Table 1

Group	N	N-scale		E-scale	
		<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>
Normals (Eysenck)	1600	0.74	—	0.75	—
Neurotics	60	0.59	< 0.05	0.54	< 0.01

regression line occur about equally on each side, but on the N-scale, the marked deviations are all on the side of relatively high short scale score with relatively low score on the remainder. The most extreme cases were three people who had the same score on the remainder as on the short scale, and one who had less, although the remainder contains three times as many questions as the short scale.

It was supposed that this effect was related to the subject's increasing awareness of the significance of the questions, with consequent concealment in some cases (it is assumed that low N scores are 'wrong' scores for neurotics), so the relationship with intelligence was considered. Fifty of the subjects had done an intelligence test (Wechsler-Bellevue I or II). Correlations between the full MPI Scales and I.Q. were as follows:

$$r_{N/I.Q.} = -0.23, \quad r_{E/I.Q.} = -0.07$$

The negative correlation between N and I.Q. is not very high (*p* is almost exactly 0.1), but it is enough to be suggestive, especially in view of the fact that in normal groups the correlation tends to be *positive* (the university students of Sigal, Star & Franks (1958) had a mean N score 3 points higher than the standardizing group). Nine subjects had N scores below 20 (the normal mean) and their mean I.Q. was 113.2, that of the remaining 41 patients being 104.0.

\* Present address: The Park Hospital for Children, Oxford

We have then two ways in which the neurotic group differs from the normal, a tendency for some subjects to give fewer neurotic responses in the later part of the test, and a tendency for some of the more intelligent subjects to give low scores. If these are both related to concealment due to awareness of the significance of the questions, awareness being increased by experience of the questions, and the more intelligent being more likely to reach such awareness, it would be expected that the negative correlation between N and I.Q. would be greater in the 'remainder' of the scale than in the short scale. These correlations were calculated, with the following result:

$$r \text{ short N/I.Q.} = -0.09, \quad r \text{ remainder N/I.Q.} = -0.26$$

This gives support to the hypothesis.

The findings of Franks, Holden & Phillips (1961) suggest that MPI scores of neurotics are less valid than those of normals in the sense that they agree less well with observation. Our findings suggest further that they are less self-consistent. On the N scale, also, this unreliability is related to intelligence. One interesting consequence of this is that for the more intelligent neurotics the short scale would appear to be a better measure of N than the full scale. We may compare this with the finding of Walton & Mather (1962) that the MMQ was a better measure for neurotics than the MPI N-scale, since its more 'physical' questions produced less 'defensiveness'.

#### REFERENCES

- CROOKES, T. G. & HUTT, S. J. (1963). The scores of psychotic patients on the Maudsley Personality Inventory. *J. cons. Psychol.* **27**, 243-247.
- EYSENCK, H. J. (1959). *Manual of the Maudsley Personality Inventory*, London: University Press.
- FRANKS, C. M., HOLDEN, E. A. & PHILLIPS, M. (1961). Eysenck's 'Stratification' theory and the questionnaire method of measuring personality. *J. clin. Psychol.* **17**, 248-253.
- SIGAL, J. J., STAR, K. H. & FRANKS, C. M. (1958). Hysterics and dysthymics as criterion groups in the study of introversion-extroversion. *J. abnorm. (soc.) Psychol.* **57**, 143-148.
- WALTON, D. & MATHER, M. D. (1962). Differential response to questionnaire items of neuroticism by 'defensive' and 'non-defensive' subjects. *J. ment. Sci.* **108**, 501-504.

*Manuscript received 29 July 1963*



## The Effects of Distributed Practice on the Learning Performance of Brain Damaged Patients

By A. E. MAY, B. R. LYND AND P. LEY

*Rainhill Hospital, Liverpool*

It has been demonstrated by Walton & Black (1957) that brain damaged patients do worse than other groups on an auditory learning test.

Two main hypotheses can account for this finding: (i) that the capacity to form the physiological connections necessary for learning has been destroyed; (ii) that brain damage has adverse effects on learning because of its effect on one or more of the variables in the learning equation.

Working with Hullian concepts Eysenck (1957) has hypothesized that brain damage leads to an increased tendency to build up reactive inhibition.

The aim of the present experiment was to see whether brain damaged patients would show improved learning in circumstances in which reactive inhibition would be expected to dissipate.

### SAMPLE

The sample consisted of 24 patients diagnosed as having some form of brain damage, and failing to learn the meaning of six out of ten new words after five presentations of the meanings.

The group ranged in age from 21 to 80 years, and had a mean age of 44.63 years, S.D. 18.23.

The diagnoses were made from clinical assessments supplemented by EEG, and psychological test finding. These were as follows:

Head Injury	6	Evidence of Brain Damage	
Presenile Dementia	4	(unspecified cause)	3
Arteriosclerosis	4	Senile Dementia	2
Intracranial haemorrhage from		Frontal Lobe Tumour	1
an aneurism	1	Prefrontal Leucotomy	1
Cerebrovascular accident	1	Narcolepsy	1
			—
		Total	24

### METHOD

Each patient was given two different sets of ten new words, on two different occasions. On one of the occasions, as a control, the administration was the usual one described by Walton & Black (1957). On the other occasion the patient was told the definition of ten new words, then asked their meanings, and then given 2 minutes rest before the next presentation of the ten definitions. This was the Experimental condition.

There were five presentations of the definitions on each occasion. No more presentations were given. Patients who do not learn the meanings by the time these have been presented five times fall outside the range of scores for the non-organic, non-defective groups in the standardization population, described by Walton and Black.

Half of the group was tested under the control condition first, and half under the experimental condition first. All patients were tested under the two conditions.

### RESULTS

Figure 1 shows the average number of words recalled on each trial. Although the average number of words recalled on the first trial of the experimental condition is higher than that for the control condition, the difference does not reach significance ( $t = 1.879$ ).

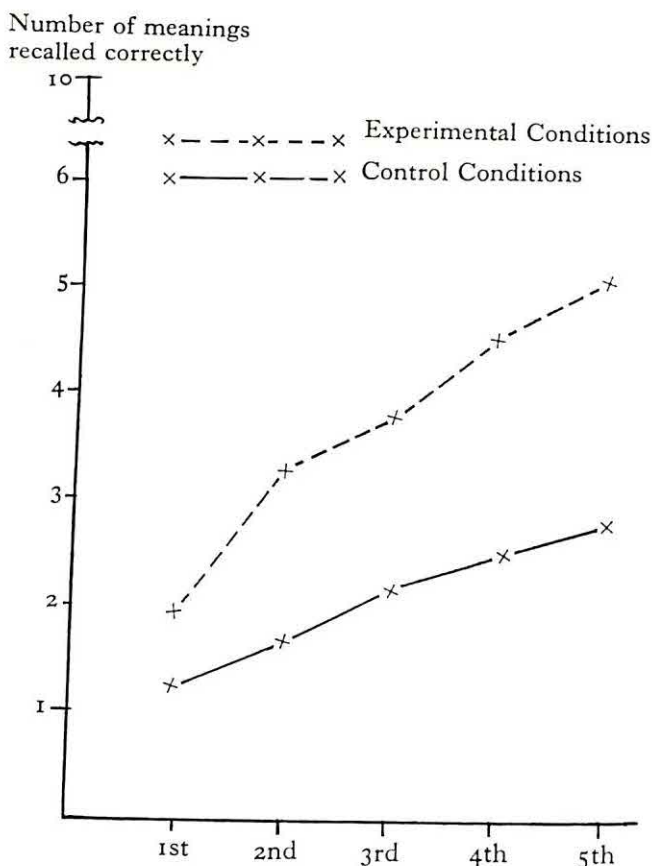


FIG. 1

The difference in total learning shown under the two conditions is significant in favour of the experimental condition ( $t = 3.346$ ).

### DISCUSSION

The results obtained lend some support to the hypothesis that the learning of brain damaged patients is impaired by excessive reactive inhibition. However, seven of the patients failed to improve under the experimental condition. This could be due to the rest pause not being long enough, for reactive inhibition to dissipate, or to an impairment of learning capacity.

Few conclusions can be drawn from the present study. It has demonstrated one condition under which the learning performance of organics can be improved and the results are not inconsistent with the hypothesis that brain damage leads to an increased tendency to build up reactive inhibition.

The authors would like to thank Dr B Finkleman and Dr J. T. Rose for granting facilities to perform this research.

### REFERENCES

- EYSENCK, H. J. (1957). *The Dynamics of Anxiety and Hysteria*. London: Routledge & Kegan Paul.
- WALTON, D. & BLACK, D. A. (1957). The validity of a psychological test of brain damage. *Brit. J. med. Psychol.* **30**, 270-279.



## The Personality of Judges as a Factor in the Validity of Their Judgments of Extraversion-Introversion

By SYBIL B. G. EYSENCK AND HANS J. EYSENCK\*

*University of London*

The questionnaire responses of nominated extraverts and introverts on the extraversion and neuroticism scales of the Eysenck Personality Inventory were studied, in conjunction with the E and N scores of the judges, and their intelligence test scores. High validity of choice was observed, but no relationship found between personality or intelligence of the judges, and the excellence of their judgments.

Personality is most frequently described in terms of behaviour patterns, and these in turn are most frequently indexed in terms of (a) self ratings, as on questionnaires, or (b) ratings by others. Both methods are open to criticism, but these criticisms are different in character; accordingly agreement between the two methods would argue in favour of the validity of both (Eysenck, 1960). Two recent studies have used the method of nominated groups (Eysenck, 1954) to test the validity of ratings of Extraversion and of Neuroticism, by having judges nominate persons supposedly extremely high or low on either of these dimensions of personality; personality inventories were then administered to these nominees, and their scores on relevant scales compared. S. B. G. Eysenck (1962) and Eysenck and Eysenck (1963a) both found evidence of considerable validity in studies using relatively small numbers of judges; they also found some presumptive evidence that some judges were better able than others to nominate persons correctly for the categories in question. The present study presents a repetition of the former experiments, with a much larger sample of judges, and extends the argument by attempting to relate the intelligence and personality of the judges to their success in judging the extraversion or introversion of their nominees. The influence of 'desirability' and 'acquiescence' response sets having been shown in earlier researches to be of relatively little import in relation to the questions used in the EPI (Eysenck, 1962; Eysenck & Eysenck, 1963b, 1963c) no special measurement was undertaken to assess their influence on our results.

The test of personality used was the Eysenck Personality Inventory (Eysenck & Eysenck, 1963), an improved version of the Maudsley Personality Inventory (Eysenck, 1959). This test has two parallel forms, but we shall here only be concerned with the combined scores from both forms. The measure of intelligence used was a well-standardized British test; this was administered in person to candidates who applied to become members of an organization (Mensa) which made the possession of a high I.Q. the prime requisite of membership. Candidates were first required to complete Form A of the test under unsupervised conditions; only those who succeeded were then admitted to the supervised test (Form B). Testing was carried out by the organization, not by the present writers, but appears to have been done conscientiously and well. From the results, two groups were formed which differed in intelligence, as defined by the test chosen. The intelligent group, with I.Q.s

\* We are indebted to the Royal Bethlem and Maudsley Hospital Research Fund for financial support, to 'Mensa' for co-operation in securing subjects, and to Mr A. Hendrickson for preparing the computer programs.

above 148 on this test, will be denoted *M* in this study, and was made up of individuals who passed the test; the less intelligent group, with I.Q.s below 148 on this test, will be denoted *P* in this study, and was made up of individuals who failed the test. (The S.D. of this test being unusually high, the tested I.Q. of 148 corresponds roughly to one of 130 on the Binet or the Wechsler scale.) Names of members of both groups were kindly furnished us by the secretary of 'Mensa'. These two groups constitute the judges; they were circulated with the EPI, and invited to take part in the general scheme of research (which was not at this stage specified). Out of about 1500 *M*-group members, 751 filled in the original questionnaire; out of 317 *P*-group members, 229 did. Details regarding the E and N scores of these subjects are given in Table 1, together with the scores for the EPI standardization group of 1931 (which, of course, did not contain either *M* or *P* members).

Table 1. *E and N scores of M and P groups, as compared with normal standardization sample*

	E		N		<i>n</i>
	<i>M</i>	$\sigma$	<i>M</i>	$\sigma$	
<i>M</i>	20.213	7.541	17.177	8.985	751
<i>P</i>	22.699	7.709	18.432	8.840	229
Control group	26.264	7.742	19.557	9.038	1931

It will be seen that apart from being more intelligent than the general population, the *M* group, and to a lesser extent the *P* group, is slightly less neurotic and much less extraverted. The former may be a reflection of the preponderance of middle-class members in both *M* and *P* (S. B. G. Eysenck, 1960; Eysenck, 1964); the latter is possibly a function of the rather cognitively-oriented type of society to which subjects belonged, or aspired to belong. Ninety-two *M* and 27 *P* members were retested about 1 year later in person when they came to the Institute of Psychiatry in order to carry out some personality tests; the retest reliability for E and N was found to be 0.88 and 0.84 for *M* members, and 0.94 and 0.92 for *P* members. The correlation between E and N for the standardization group was -0.04; for the *M* and *P* groups it was -0.24 and -0.15.

*M* and *P* members were asked to act as 'judges' or selectors, and to choose one extreme extravert and one extreme introvert each from among their acquaintances. They were furnished with descriptions of 'typical' extraverts and introverts, as follows:

'The typical extravert is sociable, likes parties, has many friends, needs to have people to talk to, and does not like reading or studying by himself. He craves excitement, takes chances, often sticks his neck out, acts on the spur of the moment, and is generally an impulsive individual. He is fond of practical jokes, always has a ready answer, and generally likes change; he is care-free, easy-going, optimistic, and likes to "laugh and be merry". He prefers to keep moving and doing things, tends to be aggressive and lose his temper quickly; altogether his feelings are not kept under tight control, and he is not always a reliable person.'

'The typical introvert is a quiet, retiring sort of person, introspective, fond of books rather than people; he is reserved and distant except to intimate friends. He tends to plan ahead, "looks before he leaps", and distrusts the impulse of the moment. He does not like excitement, takes matters of everyday life with proper seriousness, and likes a well-ordered mode of life. He keeps his feelings under close control, seldom behaves in an aggressive manner, and does not lose his temper easily. He is reliable, somewhat pessimistic, and places great value on ethical standards.'

From nominations made by *M* and *P* members, 302 and 92 replies respectively were received from nominated extraverts, and 335 and 88 replies respectively from nominated introverts. The mean E and N scores of these groups are shown in Table 2. It will be seen that the nominated extraverts have E scores of 31, while the nominated introverts have E scores of 16, i.e. almost exactly one-half as large. Both differ significantly from the population mean of 26, the introverts more so than the



extraverts. On N the nominated extraverts have lower scores than the nominated introverts, but the difference is slight (18 as against 20). This may be compared with the population mean of 20. It is apparent that, as in the previous studies, judges have no difficulty in identifying individuals who are extreme in extraversion or introversion, and it is also apparent that in doing so they do not fall into the error of confounding introversion and neuroticism to any considerable degree. The more intelligent *M*-group members do not judge extraversion better than the less intelligent *P*-group members. This argues against I.Q. as an important element in judging personality, although at lower levels it may of course exert a stronger influence.

Table 2. Mean *E* and *N* scores of nominated extraverts and introverts

	E		N			<i>n</i>
	M	$\sigma$	M	$\sigma$		
<i>M</i>	31.106	6.702	17.215	9.129	Extraverts	302
<i>P</i>	31.773	6.758	18.761	9.054		92
<i>M</i>	16.030	6.968	19.812	9.333	Introverts	335
<i>P</i>	15.924	6.064	19.739	10.284		88

Among the nominees discussed above, many had no partners; i.e. some judges nominated an extravert who forwarded his questionnaire to us, but either failed to nominate an introvert, or nominated one who failed to forward his questionnaire. Similarly, some introverts had no matching extraverts. In all, there were 225 matched pairs nominated by *M* members, and 75 matched pairs nominated by *P* members. Table 3 gives the correlations between the *E* and *N* scores of judges, and the *E* and *N* scores of nominees, separated into extraverted and introverted nominees. The argument underlying this calculation was as follows. In the group of extraverted nominees, a high *E* score constitutes a 'good' choice, while in the group of introverted nominees, a high *E* score constitutes a 'poor' choice. If extraverted judges are better (or worse) than introverted judges in making good choices, then their *E* scores should correlate positively (or negatively) with the *E* scores of their choices. A similar argument applies to the *N* scores of the nominees, although there of course both a positive or a negative correlation would indicate that judges of the particular type of personality being correlated with *N* were erroneously choosing too many (or too few) extraverts or introverts because in their minds this dimension was adulterated with *N*. The figures in Table 3 do not suggest any relationship between judges' personality and accuracy of judgment, being uniformly low. (Levels of significance required for the 5 per cent and 1 per cent level of signi-

Table 3. Correlations between *E* and *N* scores of nominated groups ( $E_{nom.}$  and  $N_{nom.}$ ) and *E* and *N* scores of judges ( $E_x$  and  $N_x$ )

	$E_{nom.} \cdot E_x$	$E_{nom.} \cdot N_x$	$N_{nom.} \cdot E_x$	$N_{nom.} \cdot N_x$		<i>n</i>
E group	0.104	-0.069	+0.028	-0.023	Mensa	225
	-0.039	-0.128	-0.048	-0.006	<i>P</i>	75
I group	0.139	0.056	-0.094	0.044	Mensa	225
	0.105	0.090	-0.146	0.221	<i>P</i>	75

ificance for the *M* and *P* groups respectively are 0.13 and 0.18 for *M*, and 0.22 and 0.29 for *P*.)

It will have been noticed that in the *M* and *P* groups, both of which were more introverted than the standardization group, *N* and *E* correlated negatively, while in the standardization group the correlation was quite negligible. These figures suggest the possibility that introverted groups in general may be characterized by a negative relationship between *E* and *N*, while the opposite may be true of extraverted groups. This hypothesis can, of course, be tested on our nominated *E* and *I* groups. The actual correlations for *M*-nominated introverts and *P*-nominated introverts were -0.19 and -0.10; those for the nominated extravert groups were -0.06 and +0.01. The evidence is slight but significantly in favour of the existence of a negative relationship between *E* and *N* among introverts; this is in good agreement with the finding of a curvilinear regression line reported in connection with the MPI (Eysenck, 1959).

#### REFERENCES

- EYSENCK, H. J. (1954). *The Psychology of Politics*. New York: Harper.
- EYSENCK, H. J. (1959). *The Maudsley Personality Inventory*. London: University Press. San Diego: Educ. & Indust. test. serv., 1962.
- EYSENCK, H. J. (1960). *The Structure of Human Personality*. New York: Macmillan.
- EYSENCK, H. J. (1962). Response set, authoritarianism and personality questionnaires. *Brit. J. soc. clin. Psychol.* **1**, 20-24.
- EYSENCK, H. J. (1964). Smoking, personality and psychosomatic disorders. *J. psychosom. Res.* to appear.
- EYSENCK, H. J. & EYSENCK, S. B. G. (1963). *The Eysenck Personality Inventory*. San Diego: Educ. & Indust. test. serv. Also: London, University of London Press.
- EYSENCK, S. B. G. (1960). Social class, sex and response to a five part personality inventory. *Educ. psychol. Meas.* **20**, 47-54.
- EYSENCK, S. B. G. (1962). The validity of a personality questionnaire as determined by the method of nominated groups. *Life Sciences*, **1**, 13-18.
- EYSENCK, S. B. G. & EYSENCK, H. J. (1963a). The validity of questionnaire and rating assessments of extraversion and neuroticism, and their factorial stability. *Brit. J. Psychol.* **54**, 51-62.
- EYSENCK, S. B. G. & EYSENCK, H. J. (1963b). Acquiescence response set in personality questionnaires. *Life Sciences*, **2**, 144-147.
- EYSENCK, S. B. G. & EYSENCK, H. J. (1963c). An experimental investigation of 'desirability' response set in a personality questionnaire. *Life Sciences*, **2**, 343-355.

*Manuscript received 24 June 1963*



## **Intra-Individual Variability in the MMPI-CPI Common Item Pool**

By LEWIS R. GOLDBERG

*University of Oregon and Oregon Research Institute*

AND RALPH M. RUST

*San Francisco State College*

Typically, attempts to measure intra-individual variability on personality inventories have relied upon the administration and subsequent re-administration of an entire test (or series of tests). Raine & Hills (1959), using projective material (sentence completion stems), have proposed using equivalent items embedded within one testing as a way of eliminating some of the well-known fatigue and motivational constraints induced by a test-retest situation (Windle, 1954, 1955). At the present time, our knowledge of item characteristics may not permit an extension of the concept of 'directly equivalent items' to structured inventory measurement, but the fact that the Minnesota Multiphasic Personality Inventory (MMPI) and the California Psychological Inventory (CPI) share many items makes the concurrent administration of both instruments a form of test-retest for their common item pool.

This method of assessing intra-individual variability has such antecedents as Buechley & Ball's (1952) Tr-scale, composed of fourteen of the sixteen duplicated MMPI items, and the Consistency scale (Con) of the Edwards Personal Preference Schedule (EPPS), based upon fifteen duplicated EPPS items. Since the MMPI-CPI common item pool is of far greater size than typical item pools duplicated within one test, it should serve as a useful base against which to compare other and shorter indices of intra-individual variability.

The present investigation represents a preliminary attempt to (a) demonstrate the usefulness of the MMPI-CPI common item pool for measuring intra-individual variability, and (b) examine critically the two shorter duplicated item pools already provided in the MMPI and the CPI.

### **PROCEDURE**

Subjects who had taken both an MMPI and a CPI were obtained from the Stanford University Counseling and Testing Center and from an undergraduate personality assessment course. The total sample included eighty-eight males and thirty-nine females. Since subjects were selected for this study solely because test data were available for them, they no doubt do not constitute a representative sample of the group from which they were obtained. The conditions under which the tests were taken and the interval between the first and second test varied widely; consequently, generalizations from this study should be extended with caution.

Five different change scales were developed for the MMPI and the CPI, as described below.

#### *Identical items (I)*

The number of items changed on re-test from the 167 items which are identical on the

MMPI and the CPI comprised the *Ss* score on this scale. The direction of change in this and subsequent scales was not considered.

#### *Virtually identical items (V)*

The number of items changed on re-test from the eleven virtually identical items on the two tests comprised the *Ss* score on this scale. Examples of virtually identical items include: 'I sometimes . . .' on one test, 'Sometimes I . . .' on the other; ' . . . worthwhile . . .' on one test, ' . . . worth while . . .' on the other.

#### *Similar items (S)*

The number of items changed on re-test out of the thirty-five MMPI-CPI items that are similar in content (though not identically phrased) comprised the *Ss* score on this scale. Differences in 'similar' MMPI and CPI items vary from the use of a different word to describe frequency or extent (e.g., 'sometimes' vs 'at times') to such an item difference as 'Some people exaggerate their troubles in order to get sympathy' vs 'I think a great many people exaggerate their misfortunes in order to gain the sympathy and help of others'. In all such items there are easily identifiable key phrases.

#### *MMPI duplicated items (M)*

The number of items changed on re-test out of the sixteen items appearing twice in the booklet form of the MMPI comprised a *Ss* score on this scale.

#### *CPI duplicated items (C)*

The number of items changed on re-test out of the twelve duplicated items in the CPI comprised a *Ss* score on this scale.

## RESULTS

### *Properties of the five change scales*

The mean, standard deviation, range of scores, and split-half reliability coefficients for each of the five change scales are presented in Table 1. Note that the average subject changed approximately 10-15 per cent of the identically worded items, approximately the same proportion of the virtually identical items, about 20 per cent of the similarly worded items, and about 5-8 per cent of the items duplicated within either test.

An indication of the generality of the various change scales can be obtained from the intercorrelations among the five measures, as shown in Table 2. None of these correlations is high, as one might expect in view of the low internal consistency of the scales. The correlations between the I-scale and the S-scale for both men and women are significant at the 0.01 level. As Table 2 indicates, the more items in a change scale, the more reliable is the measure. For men, the I-scale (167 items) has the highest average intercorrelation with other scales, closely followed by the S-scale (35 items). For women, both the I- and S-scales have similar patterns of correlation with other scales. For both men and women, the intercorrelations of these two scales with other scales are considerably higher than the intercorrelations among the other three scales (containing 11, 12, and 16 items).

Table 1. Means, standard deviations, range of scores and split-half reliability coefficients for five MMPI-CPI change scales

Scale	Number of items	Range of scores	Males (N = 88)			Females (N = 39)		
			M	S.D.	$r_{tt}$	M	S.D.	$r_{tt}$
I	167	3-54	24.6	9.8	0.61*	20.3	7.7	0.64*
V	11	0-6	1.2	1.1	—	1.0	0.9	—
S	35	1-12	6.9	2.4	-0.09	7.1	2.7	0.24
M	16	0-5	0.8	0.8	-0.08	0.9	1.1	0.28
C	12	0-4	1.0	0.9	-0.01	1.0	0.8	-0.12

\*  $p < 0.01$ .



Nineteen Ss in the present sample had taken the EPPS. Correlations between the EPPS Consistency score (composed of fifteen duplicated items) and the various change scores ranged from 0.16 for the I-scale to -0.13 for the V-scale.

Table 2. *Intercorrelations among the five MMPI-CPI change scales*

	I	V	S	M	C
I		0.38†	0.49†	0.20	0.31†
V	0.38*		0.20	0.02	0.04
S	0.48†	0.35*		0.22*	0.18
M	0.02	0.03	0.00		0.04
C	0.15	-0.27	0.20	0.06	

Above diagonal: males (N = 88). Below diagonal: females (N = 39).

\*  $p < 0.05$ ; †  $p < 0.01$ .

## DISCUSSION

This study suggests that intra-individual variability on personality inventories can be measured with modest reliability when the item pool is sufficiently large. It would appear that change scores based on small numbers of items (such as the sixteen MMPI duplicated items, the twelve CPI duplicated items, or the fifteen-item EPPS Consistency scale) are characterized by (a) low intra-individual variability (i.e., a small range of scores), (b) low internal consistency, and (c) low intercorrelations with other measures of intra-individual change. Of the change measures employed in this study only the 167-item I-scale appears to provide a reliable measure of item change. With uncorrected split-half correlations of 0.61 for men and 0.64 for women the reliability of this scale compares favourably with the reliability coefficients of MMPI clinical scales (e.g., Gilliland & Colgin, 1951, obtained *corrected* split-half correlations for MMPI scales ranging from -0.05 to 0.81; Mdn., 0.58). It is probably safe to assume that the greater reliability of the I-scale (as compared to others from the MMPI-CPI common item pool) is due to the larger number of items in the scale rather than to the small differences in the phrasing of items in some of the other scales.

This study was carried out at the Counseling and Testing Center at Stanford University. The authors wish to acknowledge the encouragement of John Black and Harold Korn. Funds for this study were provided by the Stanford University Counseling and Testing Center, a grant by Stanford University to the senior author, and National Science Foundation Grant G-25123. The Computation Center at Stanford University supported most of the statistical analysis. An extended report of this study is available from the authors.

## REFERENCES

- BUECHLEY, R. & BALL, H. (1952). A new test of validity for the group MMPI. *J. cons. Psychol.* **16**, 299-301.
- GILLILAND, A. R. & COLGIN, R. (1951). Norms, reliability, and forms of the MMPI. *J. cons. Psychol.* **15**, 435-438.
- RAINE, W. J. & HILLS, J. R. (1959). Measuring intra-individual variability within one testing. *J. abnorm. (soc.) Psychol.* **58**, 264-266.
- WINDLE, C. (1954). Test-retest effect on personality questionnaires. *Educ. psychol. Measmt.* **14**, 617-633.
- WINDLE, C. (1955). Further studies of test-retest effect on personality questionnaires. *Educ. psychol. Measmt.* **15**, 246-253.

*Manuscript received 11 October 1963*

*Revised manuscript received 25 November 1963*

## SHORT NOTES

### Validation of the Maudsley Personality Inventory in Chile

By A. C. BOLARDOS\*

*University of Chile*

The Maudsley Personality Inventory was translated into Portuguese and applied to sixty normal and fifty-one neurotic subjects of whom nineteen were diagnosed as hysterics and thirty-two as dysthymics. The mean neuroticism score of the neurotic group was 34, s.d. 8.5; that of the normal group was 17.2 with s.d. of 8.0. The difference was significant at the  $p < 0.01$  level, and gives rise to a correlation between MPI and clinical diagnosis of 0.78. Hysterics and dysthymics had neuroticism scores of 34.5 and 32.8 respectively; the difference is not significant.

On the extraversion scale it was found that hysterics had a score of 27.8 while dysthymics had a score of 17.9, a difference significant at the  $p < 0.1$ . All the results are in agreement with predictions made from the original theory on which the MPI was based, and shows that these results can also be obtained in populations differing widely in tradition and culture from those on whom the test was originally standardised and validated.

### Some Data on the Maudsley Personality Inventory in Panjabi

By S. JALOTA†

*University of Gorakhpur*

A Panjabi version of the Maudsley Personality Inventory was prepared and administered to seventy-five male and seventy-five female students at the Panjabi University, Chandigarh. For the full scale the mean neuroticism score for the male and female groups combined was 23.2 with a s.d. of 10.0; this corresponds with English norms of 19.9, s.d. 11.0. For the extraversion scale the mean combined score was 27.8, s.d. 6.2; this compares with English norms of 24.9, s.d. 9.7. There were no differences of any significance between males and females and the data suggest that the Indian group was slightly more neurotic and extraverted than the English standardization group. Findings with the short scale are similar showing a mean neuroticism score of 7.1 (6.2 for the English group) and an extraversion score of 8.2 (8.0 for the English group). The standard deviations, here as in the long scale, are rather smaller for the Indian groups than for the English standardization group.

The correlation between N and E for the long scale was  $-0.223$  which is in good agreement with the English norms. The reliability of the neuroticism scale was 0.72, that of the extraversion scale 0.53, when scores on the first half were compared with scores on the second half of the scale. These figures are lower than the English data but it is likely that an odd/even reliability would in any case give higher figures than would a comparison of first half versus second half. The data suggest that the Panjabi version of the MPI gives results not essentially different from those obtained with the original version in England, and it seems reasonable to conclude that these two personality dimensions can be found among Indian students as well as among the various European and American groups on whom the test had been standardized.

\* This paper was sent to Professor H. J. Eysenck, Institute of Psychiatry, London, S.E.5, who translated it from the Portuguese and abbreviated it for publication.

† This paper was sent to Professor H. J. Eysenck, Institute of Psychiatry, London, S.E.5, who slightly shortened it for publication.



# Hypnotizability, Suggestibility, and Personality: IV. A Study with the Leary Interpersonal Check List

By THEODORE XENOPHON BARBER AND DAVID SMITH CALVERLEY

*Medfield Foundation, Harding, Massachusetts, U.S.A.*

A study with 249 Ss failed to demonstrate that either hypnotizability or suggestibility is related to personality characteristics as measured by the Leary Interpersonal Check List.

This paper presents a study designed to test the general hypothesis that individual differences in hypnotizability and suggestibility are related to differences among individuals in personality characteristics. The criteria instruments were: the Barber Suggestibility Scale (BSS) (Barber & Calverley, 1963*b*); and the Leary Interpersonal Check List (ICL) (Leary, 1957).

## METHOD

Subjects were 249 college undergraduates (167 females and 82 males), paid 1 dollar.† When asked to volunteer for the experiment they were told that half, selected at random, would be hypnotized, and the others would serve in a control group.

All subjects were tested individually by one experimenter (D.S.C.). About half, seventy-eight females (Group A) and forty-three males (Group B), were assessed individually on *hypnotizability* by administering the BSS immediately upon completion of a standardized 15-min hypnotic induction procedure. The remaining 128 subjects were rated on *suggestibility* by administering the BSS *without* a preceding hypnotic induction procedure. Immediately prior to assessment of suggestibility, forty-one females (Group C) and twenty-one males (Group D) were given brief task-motivating instructions, as presented verbatim by Barber & Calverley (1962, 1963*e*), and forty-eight females (Group E) and eighteen males (Group F) were not given such instructions.

After completing the above, the ICL was administered and scored as described by Bentler (1963).

## RESULTS

Pearsonian correlations were computed separately for each group between objective scores on the BSS and scores on each dimension and each factor on the ICL. Table 1 shows that two of

Table 1. *Correlations of Leary Interpersonal Check List with hypnotizability (Groups A and B) and suggestibility (Groups C-F).*

Group	A	B	C	D	E	F
Sex of subjects	Female	Male	Female	Male	Female	Male
Number of subjects	78	43	41	21	48	18
<i>ICL Dimensions</i>						
AP Managerial-Autocratic	0.14	0.12	-0.18	0.28	0.14	0.07
BC Competitive-Narcissistic	0.02	0.01	-0.19	0.37	-0.17	-0.03
DE Aggressive-Sadistic	-0.12	0.03	-0.14	0.19	0.00	-0.41
FG Rebellious-Distrustful	-0.06	-0.04	0.15	-0.02	-0.15	-0.11
HI Self Effacing-Masochistic	0.01	0.02	0.00	0.38	0.37*	0.09
JK Docile-Dependent	0.17	0.21	-0.10	0.34	-0.06	0.05
LM Cooperative-Overconventional	0.24*	0.06	0.04	0.28	0.04	-0.01
NO Responsible-Hypernormal	-0.03	0.08	-0.11	0.14	0.14	-0.12
<i>ICL Factors</i>						
I. AP + BC + DE - HI	0.01	0.06	-0.20	0.23	0.04	0.09
II. HI + JK + LM + NO	0.18	0.10	-0.04	0.32	0.19	-0.21
III. Women: DE + FG + HI + JK	-0.01		-0.03		0.08	
III. Men: BC + DE + FG + HI		0.01		0.27		0.06

\* $p < 0.05$ .

† These subjects had not participated in our other recent studies (Barber & Calverley, 1963*c*, 1963*d*, 1964) concerned with hypnotizability, suggestibility, and personality.

the sixty-six correlations were significantly different from zero at the 0.05 level. Since at least three correlations should have been significant at the 0.05 level by chance alone, we cannot safely reject the null hypothesis.\*

### COMMENT

The present study, taken together with an extensive series of previous studies (reviewed by Barber, 1964), which failed to find reliable relations between personality characteristics and hypnotizability or suggestibility, suggest the possibility that personality characteristics *per se* may not play a critical role in determining response to suggestions. If personality characteristics, at least as measured by present-day methods, do not account for the major portion of the variance, what factors are of major importance? A series of recent experiments (Barber & Calverley, 1962, 1963a, 1963e, 1963f, 1963g) indicates that individual differences in suggestibility and hypnotizability may be due primarily to factors which are more situationally variable than those commonly subsumed under the concept of personality characteristics. These factors appear to include the subject's relationship with the experimenter, the subject's attitudes and goals in the immediate test situation and his motivation to perform well or poorly on assigned tasks.

This research was supported by grants (MY4825 and MH7003) from the National Institute of Mental Health, United States Public Health Service.

### REFERENCES

- BARBER, T. X. (1964). Hypnotizability, suggestibility, and personality: V. A critical review of research findings. *Psychol. Reports*, **14**, 299-320.
- BARBER, T. X. & CALVERLEY, D. S. (1962). 'Hypnotic behavior' as a function of task motivation. *J. Psychol.* **54**, 363-389.
- BARBER, T. X. & CALVERLEY, D. S. (1963a). Experimenter's tone of voice as a variable affecting 'hypnotic-like' suggestibility. Harding, Mass.: Medfield Foundation. (Mimeo.)
- BARBER, T. X. & CALVERLEY, D. S. (1963b). 'Hypnotic-like' suggestibility in children and adults. *J. abnorm. (soc.) Psychol.* **66**, 589-597.
- BARBER, T. X. & CALVERLEY, D. S. (1963c). Hypnotizability, suggestibility, and personality: II. Assessment of previous imaginative-fantasy experiences by the As, Barber-Glass, and Shor questionnaires. *J. Clin. Psychol.*, in press.
- BARBER, T. X. & CALVERLEY, D. S. (1963d). Hypnotizability, suggestibility, and personality: I. A study with the Edwards Personal Preference Schedule, the Jourard Self-Disclosure Scale, and the Marlowe-Crowne Social Desirability Scale. Harding, Mass.: Medfield Foundation. (Mimeo.)
- BARBER, T. X. & CALVERLEY, D. S. (1963e). The relative effectiveness of task-motivating instructions and trance-induction procedure in the production of 'hypnotic-like' behaviors. *J. nerv. ment. Dis.* **137**, 107-116.
- BARBER, T. X. & CALVERLEY, D. S. (1963f). Toward a theory of 'hypnotic' behavior: Effects on suggestibility of defining the situation as hypnosis and defining response to suggestions as easy. Harding, Mass.: Medfield Foundation. (Mimeo.)
- BARBER, T. X. & CALVERLEY, D. S. (1963g). Toward a theory of hypnotic behavior: Effects on suggestibility of task motivating instructions and attitudes toward hypnosis. *J. abnorm. (soc.) Psychol.* **67**, 557-565.
- BARBER, T. X. & CALVERLEY, D. S. (1964). Hypnotizability, suggestibility, and personality: III. A study using teachers' ratings of children's characteristics. *J. Psychol.* **57**, 275-280.
- BENTLER, P. M. (1963). Trust and hypnosis. Paper presented at the Western Psychological Assoc., Santa Monica, Calif., April, 1963.
- LEARY, T. (1957). *Interpersonal Diagnosis of Personality*. New York: Ronald.

\* It should be noted, however, that the small but significant correlation ( $r = 0.24$ ) obtained for one female group between hypnotizability and the Cooperative-Overconventional dimension of the ICL is in line with findings recently presented by Bentler (1963).



## Neuroticism, Extraversion and 'Hypnotizability': A Reply

By H. B. GIBSON

*The Institute of Criminology, University of Cambridge*

Hammer (1963) sets out to test an inference which he draws from a study by Furneaux & Gibson (1961) concerning the relationships between scores on the Maudsley Personality Inventory (MPI) and susceptibility to hypnosis. His paper calls for a rejoinder on four counts: (i) his inference is quite unsound on logical grounds; (ii) he seriously misrepresents what was actually done in the study to which he refers; (iii) the results he publishes have little direct bearing on 'hypnotizability' as defined by Furneaux & Gibson; (iv) his results are an artifact predictable from the similarity of the items he adds to a questionnaire, to the content of one of its scales.

A broad generalization (which requires the qualifications stated later) from the Furneaux and Gibson study is that stable extraverts and neurotic introverts tend to be especially susceptible to hypnosis. Hammer quotes some unpublished work to the effect that certain questionnaire items tend to discriminate somnambules from non-somnambules. Writing of his own *ad hoc* questionnaire he goes on to say, 'Although the correlation of the questionnaire with hypnotizability is far from perfect, it may still be inferred that if the Furneaux-Gibson conclusion is correct, then stable extraverts and neurotic-introverts will probably score more highly on a hypnotizability questionnaire than the other groups. The present investigation tests that inference.' Thus, according to Hammer, if A correlates with B, and B correlates with C, then we may infer that A will correlate with C! Such an inference is entirely unwarranted and can only be based on a fallacious concept of the nature of correlation. A classical case of two variables which both independently correlate significantly with a third, yet do not themselves correlate together, must be well known to all who are familiar with the literature of hypnosis. It has been found that both the Body Sway Test and the Heat Illusion Test correlate significantly with hypnotic susceptibility (measured by actual hypnotic induction) but neither of these two predictive tests correlate together significantly (Eysenck & Furneaux, 1945). Since Hammer's inference is fallacious, his study is incapable of testing whether 'the Furneaux-Gibson conclusion is correct'.

From Hammer's paper, it would appear that in the Furneaux and Gibson study the MPI data are comparable to those upon which his own study was based. This is not the case, and the relationships between scales and hypnotic susceptibility would certainly not be observable in any such unselected data, for reasons made clear in the Furneaux and Gibson paper, and another paper (Gibson, 1962), which is referred to in it. It is necessary to briefly recapitulate what was done in the joint study in order to emphasize the essential difference between the MPI data used in it, and those used in Hammer's study. A selection of 77 MPI protocols was made from a total pool of 99 protocols, on the basis of the Lie scale, and it was *among this limited group of Non-liars only* that the relations between extraversion, neuroticism and hypnotic susceptibility were observed as reported. Full details were given of why these relationships could be demonstrated only in this limited group. Hammer has chosen to ignore this essential point entirely in referring to the study, which is thus misrepresented, and he has conducted his own study without any use of the Lie scale of the MPI. According to the theory postulated in the Furneaux and Gibson paper, because of this omission he would *not* have found the relationships they report even if he had used an actual test of 'hypnotizability' instead of his ten-item questionnaire. Subjects high on the Lie scale tend to distort the Neuroticism scale, and they also tend to be resistant to hypnosis.

Furneaux and Gibson claimed that sixty-four questionnaire items, used in special combinations, successfully discriminated 71 per cent of those susceptible to hypnosis in a pre-selected population. Hammer now claims that only ten items, thrown together in a new 'scale', will

assess 'hypnotizability' directly in the general population. So impressive a claim deserves careful examination.

Of these ten items, three (Nos. 55, 57 and 58) had never been used in any study before, but we are told that their wording was 'similar to' or 'resembled' the wording of items which had been used in a small unpublished study in which hypnosis had been used. Two other items (Nos. 49 and 56) had actually been used before in that unpublished study. The remaining five items were extracted from a longer scale in another unpublished study, and they had been found to discriminate between twenty-four somnambules and controls in a pre-selected population. As to the *content* of the items, they are nearly all concerned with absent-mindedness and reverie, and have a close similarity to the four MPI Neuroticism items which are also concerned with absent-mindedness and reverie. All of them would, in terms of face validity, serve as a reasonable extension to the MPI Neuroticism scale. Like that scale, they are all coded positively. What Hammer has done, is to add on ten items to the end of the MPI, and to discover what could easily have been predicted, that the score on these ten items correlates hugely ( $p < 0.000001$ !) with the score on the scale it resembles so closely. As he chooses to call a high score on the ten items 'hypnotizability', we then have the myth perpetuated that hypnotizability is a direct function of neuroticism. His study thus has no relevance at all to the issues raised in the Furneaux and Gibson study he is supposed to be testing. It would be unfortunate if the subject of hypnosis, already burdened with too many myths, should be further confused by the propagation of a misunderstanding.

#### REFERENCES

- EYSENCK, H. J. & FURNEAUX, W. D. (1945). Primary and secondary suggestibility: an experimental and statistical study. *J. exp. Psychol.* **35**, 485-503.
- FURNEAUX, W. D. & GIBSON, H. B. (1961). The Maudsley Personality Inventory as a predictor of susceptibility to hypnosis. *Int. J. clin. exper. Hypnosis.* **9**, 167-177.
- GIBSON, H. B. (1962). The Lie scale of the Maudsley Personality Inventory. *Acta psychol. Hague*, **20**, 18-23.
- HAMMER, A. G. (1963). Neuroticism, extraversion and hypnotizability. *British J. soc. clin. Psychol.* **2**, 185-191.

*Manuscript received 5 November 1963*



## Book Reviews

*Individual in Society.* By D. Krech, R. S. Crutchfield and E. L. Ballachey. New York: McGraw-Hill, 1962. Pp. 564. 62s.

Krech and Crutchfield's most recent text-book the *Individual in Society* is also their best. It is quite excellent. This text-book, written in collaboration with Ballachey, does much more than bring up to date *Theories and Problems of Social Psychology*. The new text-book is the result of very considerable rethinking on the part of the authors of what constitutes the essence of social psychology, and how social psychology links up sociology and anthropology. The *Theories and Methods* text-book used a tight conceptual framework within which the individual was seen in interaction with his environment; the extent to which this interaction was affected by the subject's reference groups was touched upon, but not developed. It is in this respect that the new text-book constitutes such an advance. Here, to use the Parsonian terminology, the social and cultural systems are introduced into the body of the book and help to show the uselessness of demarcating clearly the psychological and sociological field.

This text-book has three great virtues. First, the range of social science disciplines and studies on which the authors draw is an unusually wide one. For example, the volunteer trials to isolate the common cold virus are quoted as appropriate social psychological data, alongside the results of standard laboratory experiments. In this way, the student is shown the wider applicability of the processes he is examining. Second, the authors have been able to present the diversity of material in a clearly articulated framework; every example is strictly relevant to the theme under discussion. Third, the organization of the book is excellent; there is no attempt to sugarcoat the pill, but it is written on the assumption that the material is exciting, that there is no need to 'write down'. Finally, the text-book scores because of the various devices it adopts; the device of a glossary in which the concepts used are explained, the guides which set out the steps in the argument, and, above all, the device of inserts or boxes by which the authors manage to present an uninterrupted account of the process under discussion while at the same time offering the students a detailed account of relevant studies.

Because of the clarity of exposition and the absence of 'patronage', this text-book is suitable for three types of audience: the undergraduate who wants to acquire some understanding of what constitutes social psychology; the fully-fledged psychologist who is not sure what social psychology is like, but as often as not is convinced that it is non-rigorous and intellectually non-demanding; and the graduate student who specializes in general or social psychology.

The book covers the usual areas of social psychology, beginning with a very prolonged account of cognition and motivation, and with a very full and exhaustive treatment of attitude measurement, and theories of attitude formation. It is least informative on roles and norms, relative to the very thorough treatment given to other areas. The section dealing with groups and leadership is excellent.

It is fitting that this text-book should be dedicated to Tolman. Not only was he the teacher and friend of the three authors, but the book so clearly reflects the basic tenet of his teaching, that all connected with the human mind is of interest to the psychologist and that the psychologist, to test the universality of the processes he isolates, needs to examine these in different social and cultural settings.

HILDE HIMMELWEIT

*Six Cultures: Studies in Child Rearing.* Edited by Beatrice B. Whiting. New York: John Wiley, 1963. Pp. 1017. 95s.

This bulky volume contains mainly raw material for students of culture and personality. Hitherto, if one wanted to test hypotheses in this sphere, it was necessary to rely directly or indirectly on anthropological monographs designed for a different purpose. While adult



behaviour is usually comprehensively covered, data on child rearing tend to be patchy and uneven, rendering systematic comparisons difficult. A grant from the Ford Foundation enabled scholars from Cornell, Yale and Harvard to plan a comprehensive survey of child-rearing-within-its-setting in a variety of contrasting cultures. The present volume is the first outcome of these studies.

The fieldwork was carried out by (mostly husband-wife) teams of investigators in small communities in Kenya, India, Okinawa, Mexico, the Philippines and U.S.A. A majority of the team-members were trained in social anthropology, and the fieldwork lasted between 6 and 14 months. Each report is organized, with only minor variations, along the following lines: the first part consists of an account of the broad cultural background—economy, social and political structure, marriage and kinship, religion and the supernatural, disease and death; the second part describes the main stages of child-rearing from pregnancy to early adolescence; both parts receive approximately equal weight.

The authors clearly had a difficult task, particularly when trying to provide a concise outline of the salient features of whole culture. In general it was extremely well done. The studies are readable and convey a vivid sense of the varieties of styles of life depicted. Social change is not neglected, and the curious blend of old and new is often well portrayed. For instance, there is a description of 'Mother's Day' in Okinawa, where nature deities continue to be worshipped.

The child-rearing sections are, on the whole, more uniform in presentation; this is hardly surprising, seeing they are based primarily on intensive studies of twenty-four mothers within each culture, combined with direct observation of the children's behaviour. The question of how the mother handles the child, which is a prominent and almost exclusive concern in conventional questionnaire studies of child-rearing practices, is only one of several aspects dealt with; much attention is devoted to the nature of the child's interaction with others in particular situations, and this emphasis is as valuable as it is rare. It became feasible because most of the communities are located in warm climates where a lot of activity takes place out of doors; and this is perhaps one of the reasons why the one conspicuous exception, Orchard Town, New England, seems comparatively dull. Apart from this, and the fact that the material is familiar, the section is indubitably of a lower standard; for instance, kinship and marriage are covered in less than two pages, there being no mention of sex, and one is told that 'all marriages are not considered equally preferable'. However, this strikes one so forcibly only because of the high level of excellence which characterizes most other contributions.

Nothing has been said here about the theoretical considerations which governed the design of the studies and the selection of topics. They are expounded by the editor in an introduction occupying little more than 1 per cent of the total length. Nevertheless, there would be plenty to discuss; but perhaps it is better to wait, in view of the promise of theoretical volumes yet to come. Meantime the reader is invited by the editor to try his hand at some do-it-yourself, which is rather a lot to expect. One omission about which one has the right to complain concerns methods; e.g. on what basis were the twenty-four mothers selected? These gaps will no doubt be filled in the further series of publications, to which one looks forward eagerly. The present volume is a source book teachers will find very useful, for its broader ethnographic accounts as well as for the descriptions illustrating the wide range of existing child-rearing practices.

GUSTAV JAHODA

*The Mentally Retarded Child.* Edited by A. R. Luria. Oxford: Pergamon Press, 1963. Pp. viii + 207. 50s.

Though I greatly admire Professor Luria for his brilliant and original contributions to the study of mental subnormality, I find myself unable to extend this admiration to all of the present book. Luria begins with an introductory chapter in which he gives a rather distorted picture of the views held on the subject outside the Soviet Union. Polemical statements such as that 'in capitalist countries children are not held to depend on the conditions of the environment but are determined from within by given mental capacities' are not solely representative of work in this field outside Russia. As Luria is surely well acquainted with western experimental work in mental deficiency it may well be asked why a more balanced account of such



work should not have been given. In the same chapter it is asserted rather than demonstrated that all subnormality is the result of under-development or mal-development of the brain. While one would certainly agree with this as far as severe subnormality is concerned, this case has been less well established for feeble-minded children, and Luria presents no additional evidence concerning this group. In a chapter on clinical characteristics Pevsner says that there are reasons for believing that the cause of feeble-mindedness is a diffuse and mainly superficial damage to the cerebral hemisphere. However, we are simply asked to share this belief and no relevant data is presented. This chapter by Pevsner represents an interesting attempt at sub-classification of subnormality on a functional basis, though it is couched in rather irritating pseudo-physiological jargon.

A chapter by Zislina on E.E.G. investigations of subnormal children makes again rather sweeping statements, for instance that the alpha rhythm is absent in the E.E.G.s of such children. This is not confirmed by studies carried out here and in the U.S.A., though of course abnormalities in the E.E.G. of severely subnormals are usually found.

However, from here onwards the book improves and chapters in the second half report the characteristically ingenious experiments on the role of language in the organization of behaviour of mentally subnormal children. Though this account by Luria, Vinogradov, Lubovskii and Meshcheryakov lack detailed presentation of data, and much of the material has been previously published in English, the chapters nevertheless give a clear account of the significant contributions which Luria and his colleagues have made to the subject. As in the case with most English editions of Russian psychology books, the standard of editing and translation leaves room for improvement. I cannot believe that Luria really regards epileptic children not only as irritable, greedy and rude but also as 'ugly'; nor that another patient was distinguished by 'motor restlessness and superfluous ability'. At a price of fifty shillings the book cannot really be recommended as a best buy.

BEATE HERMELIN

*Readings in Mathematical Psychology*, Vol. 1. Edited by R. D. Luce, R. R. Bush and E. Galanter. London: Wiley, 1963. Pp. ix + 535. 68s.

Under the title of this book one would expect a collection of classical papers which would enable the reader to go back to the sources of an established science. It is, in fact, a very uneven collection of papers reprinted from *Psychometrika*, psychological journals, and from various journals of engineering, logic etc. Some material comes from reports of U.S. Government sponsored research.

It seems to the reviewer that, alongside with some valuable material, there runs a stream of papers carrying much mathematical dead-wood and floating jetsam of characteristic fallacies.

The volume opens with a very valuable paper by Adams and Messick which attempts to furnish a test for Thurstone's successive interval scaling model and generalizations one might resort to when the test fails. The model is said to fit exactly if all the testable consequences of the author's axioms are verified. This definition raises a lot of problems which are never faced in this volume.

The fallacy that elaborate, often rather over-precise models are tested when some consequences are verified runs through this book. Very often the model is chosen to suit the mathematical techniques available—a principle which has something Parkinsonian. Thus the model of decision structure considered by Christie and Luce has no other justification than the convenience of manipulating Laplace transforms. It is not based on any neurological hypothesis or on any empirical knowledge concerning decision structure.

Authors of papers reprinted from journals of logic like Suppes and Scott do not necessarily subscribe to the fallacy that the axiomatics of mathematical postulates can clarify any questions of experimental psychology but the editors apparently do.

They seem to believe that elaborate deductions from assumptions are somehow worth more than the assumptions and they have an unlimited confidence in the possibility of the design of experiments which can test untested axioms 'en bloc'.

George A. Miller seems to think that the jargon of the engineer is better than plain English. To say that 'the amount of information' is exactly the same concept that we have talked about



for years under the name of 'variance' is an inexcusable over-simplification. His paper on the maximum likelihood estimate of the Shannon-Wiener measure of information comes from a Government sponsored report but was not published elsewhere. One wonders what could be the use of small-sample statistics with a concept which is only useful for the understanding of the economics of transmitting very long chains of signals, but has never been shown to have any other significance. The paper, incidentally, contains very long chains of symbols stretching over many pages of intermediate calculations with very small information content, and so do others.

The volume contains papers by Mosteller, Estes, Kanal and many others on learning theory. Other contributions deal with the determination of scales for stimuli, paired comparisons, and with detection of signals in the presence of noise, but there are, unfortunately, no formulae which the reviewer could have readily applied to the task of reviewing this book.

H. MOTZ

*Psychology and Religion.* By G. Stephens Spinks. London: Methuen, 1963. Pp. xv + 221. 28s.

The sub-title of this book describes it as 'an Introduction to Contemporary Views', but this is misleading, for it ranges over the whole field of the subject, from the cave paintings of prehistoric times, the rituals of primitive cultures, the Greek philosophers and the Christian fathers, up to the latest developments in psychological theory and the experiments of Aldous Huxley with mescaline. Dr Spinks has approached the subject from the viewpoint of the anthropologist and social psychologist rather than make a critical examination of the inner coherence and psychological structure of any given aspect of religion. The book both gains and loses as a consequence. As a survey of religious beliefs, with lucid comments to describe them, it is an excellent outline for beginners. To read it is to ensure a comprehensive perspective, so often lacking in more specialized books. But the breadth of treatment in what is a comparatively small book has inevitably resulted in a corresponding shallowness, however acute the insights are. To give but one example, in his discussion of Freud's writings on religion Dr Spinks does not take adequate account of the general theory of psychoanalysis and the light it throws on unconscious motivation and substitute satisfactions.

The greatest defect of the book is its lack of critical evaluation. By accepting a sociological definition of religion it excludes any criterion by which to distinguish pseudo-religion from true religion; all religions are accepted at face value and no principles are laid down by which to determine whether God, the object of religious faith and practice, has in any or all cases more than psychological reality. Inevitably, therefore, the validity of all forms of religion is assumed and the reader is left with little to guide him if he seeks to discriminate among them. These would be very serious omissions in a larger volume and they detract from the value even of this introduction, admirable as it otherwise is.

R. S. LEE

*Psychology in Relation to Medicine.* By R. M. Mowbray and T. Ferguson Rodger. London and Edinburgh: E. & S. Livingstone. Pp. vii + 399. 36s.

The lack of a suitable text-book of psychology for medical students has been felt for a long time; DPM candidates have also searched in vain for a book that meets their particular needs. The present book, written jointly by the Senior Lecturer in Clinical Psychology and the Professor of Psychological Medicine at Glasgow University, is addressed primarily to medical students, but it may also be a useful introduction for DPM students. On the whole, it fulfils its purpose admirably, and is written in a readable, almost colloquial style, though bearing a little too obviously the hallmarks of the lecture room, and of the Scottish scene.

The book gives concise and up to date accounts of the history, applications and methods of psychology and discusses at some length a number of theories and experimental approaches to instinct and emotion, including work in ethology, psychoanalysis and psychosomatic medicine.



Perhaps the most useful and relevant section is one on higher mental processes, which includes an excellent introduction to learning theories and their practical application, and strides without obvious signs of malaise from Thorndike via Hull to Skinner. About half the book is devoted to sections on Intelligence and Personality, with emphasis on clinical problems. Each section ends with a list of sources and suggested reading and at the end of the book there is a special reading list for the general reader.

Apart from a sound respect for Adolf Meyer and for scientific method, no particular orientation is apparent, and the authors are scrupulously fair to a variety of different approaches, from psychoanalysis to cybernetics. In fact, they might be accused of adopting, in places at least, a somewhat uncritical attitude to a number of complex and highly controversial subjects, as for example in their treatment of maternal deprivation, somatotyping and the work of Piaget. Their tendency to simplify and occasionally to dogmatise may give rise to some misgivings among psychologists, but can probably be justified by the need to communicate a measure of unequivocal fact to medical students, who, as the authors point out, are trained to deal with facts and are 'puzzled by the psychologist's insistence on knowing how a particular fact of human behaviour was derived'. Although some degree of over-simplification may be justified in the interests of communication with a potentially prejudiced and possibly hostile audience, the authors are cautious and circumspect in some parts of the book, notably in the chapter on personality assessment, where their lack of any clear orientation is particularly apparent. After a lucid account of the methodological problems and the main techniques of personality assessment, objective tests are criticized for poor reliability and validity, but projective tests are mentioned in a favourable light, and a Rorschach psychogram illustrated and interpreted in some detail. Eysenck's theoretical and clinical contributions are praised, but the MPI is rather summarily dismissed as clinically insensitive. To redress the balance, however, two pages are devoted to pictorial illustration of 'items which appear in intelligence tests', taken from Eysenck's *Know Your Own IQ*.

In discussing mental subnormality, the traditional definitions of imbecility are given, stressing the imbecile's social and educational incompetence, though a few pages later the work of O'Connor and Tizard is reported which flatly contradicts these views. The discussion on mental deficiency would also have been improved if the question of I.Q. constancy, which is here treated purely psychometrically, could have included a reference to the Clarkes' work on I.Q. rises after recovery from social deprivation.

These and similar criticisms do not seriously detract from the value of a book which should go far towards convincing medical students, and perhaps not a few of their teachers, that psychology deserves a larger share of the curriculum than it has previously received. Perhaps future editions could dispense with the profusion of exclamation marks and also bear in mind that the book is likely to be read south of the border.

P. MITTLER

*Current Perspectives in Social Psychology: Readings with Commentary.* Edited by E. P. Hollander and R. G. Hunt. New York: Oxford Univ. Press, 1963. Pp. xi + 557. 36s.

At a time when books of readings are proliferating, the present volume breaks relatively new ground in that it seeks to be representative of contemporary theoretical viewpoints in social psychology. With a fairly clear field, the authors have not hesitated to go to the obvious sources; nearly half the fifty extracts are taken from recent issues of *Psychological Review*, *American Psychologist* and *Public Opinion Quarterly*. Hence, most of them are familiar, meaty and worth reprinting. Unfortunately, and as usual, many are abridged, so that the serious student is not spared the labour of seeking out the original; and again, as usual, the excerpts from books either titillate, or irritate, or enable the reader to add further volumes to the list of books he is determined not to read.

The contents are, of course, varied. We have papers of armchair, conceptual analysis, such as those by Shibutani on reference groups and White on culture, on the one hand, and on the other, the more stimulating, semi-formal, experiment-related theories of Rokeach, Osgood,



Festinger and Kelman. Stylistically they vary from the lucidity of G. W. Allport and of McClelland to the murky complexities of Heider and of Goffman. Some, a minority, are mainly empirical with a minimum of theoretical generalization, such as Crutchfield's chatty account of his experiments on conformity; in others, reference to the empirical is brief and anecdotal. The editorial commentary consists mainly of uncritical summaries of the papers, and serves merely as a device for tying them together loosely in bundles. As a consequence the unsophisticated reader is apt to be plunged into the middle of controversies with very little preparation.

The volume, as a whole, is neither a text-book nor a book of illustrative readings to accompany a text-book. Hence it will only serve the undergraduate as an occasional reference. For the more sophisticated it offers a fairly representative bird's-eye view of the theoretical confusion in this field.

D. S. WRIGHT

*Organizational Choice.* By E. L. Trist, G. W. Higgin, H. Murray and A. B. Pollock. London: Tavistock Publications, 1963. Pp. 332. 48s.

Any thoughtful explorer during the nineteen-forties who went seeking ideas on the theory of industrial organization soon found himself in a stark and unpromising landscape. There was the Weberian model of managerial bureaucracy, authoritarian and monolithic; assumed to be intrinsic to the nature of modern industry. For the rank and file labour force, there was the unsparing logic of extreme individualistic division of labour, expressed in such phrases as 'maximal task breakdown' and 'repetitive task components'. For those excessively oppressed by this, or unconvinced that it offered the most fruitful framework for human co-operation, there was an optional extra called 'human relations'. Managers and supervisors could be trained in 'social skills' which would enable them to release the appropriate motivations in their subordinates and thereby remedy the many failures of co-operation suffered by the system.

In the past decade this bleak terrain has been transformed. Recent writings have shown that for some types of technology and market situation, the bureaucratic model is not necessarily the optimum form of management organization. An ever-increasing sub-division of labour is not necessarily, French sociologists in particular have argued, the ideal aim even for those concerned solely with efficiency. And even more fundamentally, the notion implicit in the earlier 'human relations' approach that social and psychological factors can usefully be studied in isolation from technology and organization of work is now seen as naïveté.

The importance of this book is that it brings powerful strength to all these new and stimulating lines of thought. The central concept is that of the 'socio-technical system' (which the Institute itself pioneered), with its basic premise that motivation, group structure, interaction processes and authority must all be studied in their own context of technology and work organization. The authors convincingly illustrate, from field research in coalmining, the feasibility of creating self-regulating primary work groups engaged upon integrated tasks, with supervisors pursuing a 'service' rather than a 'disciplinary' function. In so doing they show that technology does not always uniquely determine work organization, but may only set limits within which we may be able to choose between alternative forms, some of them more satisfying and productive than others. This is without question an exciting book which should soon rank as a classic in the study of the work group.

ALAN FOX

*Research Readings in Child Psychology.* Edited by David S. Palermo and Lewis P. Lipsitt. New York: Holt, Rinehart and Winston, 1963. Pp. viii + 572, \$7.50.

This collection of research studies and the methodological discussions which introduce them should prove useful for advanced courses of Child Psychology. The bias is towards experimental investigations of the kind which relate responses to variations in the stimulus conditions rather than towards studies correlating classes of behaviour. Experiments using



instrumental procedures and the operant conditioning paradigm predominate, but some longitudinal, comparative and cross-cultural studies are reported.

Any orderly arrangement of studies in a field as diverse in its theoretical and practical origins as Child Psychology presents difficulties. The editors have tackled the problem by grouping investigations around research methods rather than subject matter. At the same time, there is an attempt to treat certain classes of behaviour, like sensation and perception, discrimination, concept formation and language, together, and to make some reference to the historical and theoretical background. The division is not entirely successful. In some cases, experimental studies of essentially similar questions are assigned to different sections of the book because of the theoretical framework within which they were originally conceived. The few investigations on the development of emotional and social behaviour which are reported are scattered. Possibly a separate section on the subject might have to include some descriptive, comparative and correlational studies, and so look regrettably untidy. But it need not be confined to these, and the subject is sufficiently important to warrant adequate representation in a basic text. However, these are relatively minor criticisms of an otherwise excellent selection.

SUSANNA MILLAR

*The Sociological Review Monograph No. 7: Sociological Studies in British University Education.* Edited by P. Halmes. Univ. of Keele, 1963. Pp. 204. 32s.

This valuable and well-presented monograph consists of fourteen articles on the expansion of British Universities and the effect of this expansion on the selection of university students. Some writers are concerned with what is described on line 1 as 'the corridors of power', and their essays have more interest for sociologists than for psychologists, although they do reveal the extraordinary fact (surely not in the public interest) that three British Universities did not participate in the survey of applicants for universities conducted on behalf of the Committee of Vice-Chancellors and Principals. Several interesting contributions are made by psychologists, including Dale, Furneaux, Himmelweit and Vernon. Vernon heavily criticizes many cherished beliefs in this field. He doubts whether there exists in the population a fixed distribution, or 'pool', of intelligence. He states categorically that no calculations of the number of eligible students can be based on tests of intelligence or other aptitudes, though they could conceivably be based on tests or surveys of aspirations, interests and social attitudes in the population. In his view, the more important factors are the cultural and vocational attitudes (not merely socio-economic status) of the home and the community generally, and the individual's own motives and interests. Acceptance of these principles does not, he thinks, imply a lowering of standard, nor increased wastage. Many psychologists would now take the view that, although the genes set some limits on intellectual capacities (and it is vital that we should not dig our heads in the sand about this), our present methods of education do not exploit them to the full. It is becoming increasingly obvious that the ability of underprivileged people in Britain to rise to their opportunities has been underestimated. Furnaux suggests that this social revolution in so far as it has occurred at all, may now have lost its momentum, and that it does not appear that differences in University performance result mainly from financial determinants, but that they indicate an absence of adequate educational goals and motivation in the majority of families. Dale emphasized 10 years ago that the principal reason for failure in university examinations was not lack of intelligence but lack of application to work, although he now describes later researches which suggest home background as the main determining cause of deterioration in academic performance, and of early leaving. Himmelweit presents her well known researches suggesting that the failure rate at L.S.E. would have been reduced from 23 per cent to 8 per cent by the use of suitable specially designed tests for measuring high-grade intelligence. But, remarkably, no one has seemed to care, and Kelsall mentions, in true British fashion, the 'dangers' of efficient selection.

Although it is not possible in a short review to do even scant justice to the individual authors of an omnibus, the reader is recommended to Little's demolitionary article exploding some myths of University expansion. The rapid increase of University education outside and inside this country is clearly affecting our whole outlook on selection. With the ratio of the general



population enjoying higher education in California approaching 50 per cent, one wonders whether University placement would not be a better social objective than University selection. Moreover, a lowering of selection standards would increase the number, although not the proportion, of good students. We hear too much from our Canutians of the low standards of poor students in the great American universities, and not enough of the high standards of the good ones. This book should do a great deal to get our views on a more realistic basis.

F. W. WARBURTON

*Committees—How They Work and How to Work Them.* By E. Anstey. London: Allen & Unwin. 1962. Pp. 116. 15s.

No doubt the family is the most universal of small groups, but in modern society the committee seems to be running it pretty close. Yet there are surprisingly few good books on committees. Dr Edgar Anstey, Senior Principal Psychologist in the Ministry of Defence and unshaken veteran of many gruelling committee meetings, sets out to provide a brief introduction for committee members to the main kinds of committee, their most common failings, and ways in which improvements could be made. The result is a modest, readable and distinctly useful book.

After opening with a definition of committees which stresses their definite composition, responsibility to a parent body or bodies, terms of reference, and their need both to discuss and to decide, Dr Anstey offers a simple classification of committees according to their length of life, their purposes and their variability of composition. He then goes on to discuss the relative advantages of formal and informal methods of conducting business, linking these with the conditions under which different committees function. Much of the remainder of the book is concerned with examples of ineffective discussion and decision-making, together with liberal advice that would undoubtedly improve practice immeasurably for those free to act on it. Since many committee members are inadequately prepared for their roles, separate chapters are given to the role of the chairman, the secretary's duties, advice to a new member, and to training in committee work. A substantial appendix provides a detailed account of a specimen committee meeting used by the author in his own training courses. The book is very brief and is intended to be as practical as possible, so that it is understandable that little impression could be given of the deeper complexities of group interaction and personal motivation. But it is far more serious than the light-weight presentation, with Punch-type illustrations by Thelwell, might suggest.

J. F. MORRIS

*A Study of Children's Thinking.* By Margaret Donaldson in collaboration with Donald Withrington, with an appendix by John Duthie. London: Tavistock Publications, 1963. Pp. viii + 263. 35s.

The aim of the study reported here is to investigate the basis for the predictive power of verbal reasoning test items by tracing the kind of error made at various ages in reaching solutions. The author is fully aware that the 'thinking aloud' method she uses is exploratory, and employs it to derive hypotheses to be tested later. An intelligent analysis of the variables involved in each type of problem enables her to vary the tasks with respect to the nature of the objects or symbols to be related, the number and complexity of items, the form in which information is given, and so forth. Although this is not fully explored, her results indicate some interesting relationships between these variables and the types of error she finds. The experimental investigation to which this study is the prelude should prove rewarding, especially if it explores these relationships systematically in its search for better tests. The results are discussed in relation to a variety of other studies, but there is no comprehensive survey of the field. Perhaps the investigation would have been better published as a monograph than in book form, but it makes a worthwhile contribution to the subject.

SUSANNA MILLAR



## Psychological Assessments of Brain Damage and Intellectual Impairment in Psychiatric Patients

BY J. E. ORME, D. LEE AND M. R. SMITH

*Middlewood Hospital, Sheffield*

This study reports the apparent capacity of certain test performances to detect brain damage. Two measures (derived from the Modified Word Learning Test and from design recall) were used which have been claimed to be specific indicators of brain damage. A measure of intellectual impairment of a more general kind was derived from verbal/non-verbal intellectual performance discrepancies. A group of brain damaged subjects, three groups of neurotics, three groups of psychotics and an epileptic group were studied. The results indicate the differentiation of brain damaged subjects from others is possible. Various score combinations are discussed to this end. Other points of interest concern the performances of chronic schizophrenics, and positional learning effects in different clinical groups.

The clinical psychologist is frequently asked to help in the differential diagnosis of brain damage. The problem bristles with many difficulties, some of them at present insuperable. Certain of these may be listed. (1) Difficulties in the validation of the occurrence and extent of brain damage. (2) Individual variations in the behavioural manifestations of brain damage. (3) The role of organic factors in 'functional' disturbances and therefore in their test performances.

Yates (1954) drew attention to further difficulties in the use of test material not properly standardized for age or intellectual level. Furthermore, tests which clearly distinguish between organics and non-psychiatric controls might be useless for distinguishing between organics and psychotics.

The present report concerns a study using three techniques of assessment.

(1) The Modified Word Learning Test (Walton & Black, 1957, 1959; Walton, 1958; Walton *et al.*, 1959; Walton & Mather, 1961) has been reported as distinguishing brain-damaged patients from others with only a small amount of overlap. The subject attempts to learn the meanings of ten words, new to him, determined by vocabulary performance. Non brain-damaged subjects tend to learn to the criterion of at least six new words in less than six successive trials. Age and intellectual variations have no significant effect on the test's discriminatory value. In the present study, the words were derived from the Mill Hill Vocabulary Scale. Different vocabulary scales apparently produce different results when used for the Modified Word Learning Test (Riddell, 1962). One possible advantage to using the Mill Hill Vocabulary Scale is that it has been standardized on British populations.

(2) The second technique derives from design recall. Six of the Bender figures (A. 4, 5, 6, 7, 8) are individually exposed to the subject for 5 sec. He then has to draw the design. It has been reported (Orme, 1962) that rotation of designs, separation or omission of parts of the designs, are not specifically related to brain damage. It was claimed, however, that distortion of the actual line of the designs when reproduced was highly associated with brain damage. Each design was marked for the presence (1 point) or marked degree (2 points) of distortion. A score of 4 or more was found to be highly indicative of brain damage. Age and intellectual variations had no significant effect on the test's discriminatory value. Examples of distortion are given in the paper by Orme (1962).

(3) The third technique is a measure of verbal/non-verbal intellectual performance discrepancy. Verbal test performances (particularly vocabulary scores) are more impervious to intellectual impairment (from any source) than non-verbal items. There are many reservations to such a belief (Yates, 1956) and vocabulary itself may well be effected in organic states (Orme, 1957). Nevertheless, a verbal/non-verbal discrepancy in favour of the former is significantly associated with brain damage (Orme, 1962). In the present study, the Mill Hill Vocabulary Scale (oral definitions) and the Coloured Progressive Matrices were used as measures of verbal and non-verbal performances respectively. The Coloured Progressive Matrices (Raven, 1958) consists of thirty-six problems specially designed to give a broad scoring spectrum to the lower ranges of non-verbal performance. Table A of the Appendix gives the expected raw score difference between vocabulary and matrices performance for each age range and vocabulary level.

This table was constructed from smoothed out curves of vocabulary and matrices equivalent scores derived from the tests' published norms and further material, partly unpublished, by Orme (1961). The discrepancy estimate is obtained by subtracting the expected raw score discrepancy from the actual raw score discrepancy. Discrepancies showing a relatively high verbal performance will therefore have plus values.

The subjects used were psychiatric patients grouped according to the diagnosis made by the psychiatrist concerned. A number of psychiatrists were involved which helped to minimize any diagnostic bias possibly present in individual psychiatrists. Where there was any chance of test results contaminating diagnosis (re brain damage), care was taken to record the diagnosis made before or considerably later (up to 2 years), than when the results were communicated. Table 1 gives details of the sample.

Table 1. *Composition of sample studied*

Group	N	Age		MHVS level (percentiles)		
		Mean	S.D.	75 and above	26- 74	6-25
Organics	53	56.5	11.8	3	26	24
Anxiety states	28	44.4	11.5	0	13	15
Neurotic						
Depressives						
Hysterics	20	36.2	17.6	3	6	11
Psychopaths	30	21.4	12.6	0	16	14
Melancholics	30	51.7	26.2	8	15	7
Acute schizophrenics	24	39.8	18.5	4	7	13
Chronic schizophrenics	23	52.8	10.7	1	10	12
Epileptics	15	35.0	14.1	2	4	9

No patients were included with a vocabulary level within the bottom 5 per cent of the population. There was no evidence that varying vocabulary level above this had any bearing on the results to be described. The organic group included any patients considered to be suffering from some degree of generalized brain damage (usually cortical), and included senile dementias, head injuries, brain tumours, Korsakoff states etc. There are obvious age differences following the pattern usually seen between the various clinical groups. Variations in age were not found to affect the results in any way.

## RESULTS AND DISCUSSION

Discussion of the significance of the findings of the groups of chronic schizophrenics and epileptics will be left till later. Concerning the other groups it must be stressed that the major object of this study was the examination of certain techniques as practical detectors of brain damage. That is, statistically significant differences of



only theoretical interest are not sought. It is rather the evidence for a difference in scores between organic and other groups so great that score overlap is minimal. Accordingly, the cut-off points established in previous publications on the Word Learning Test and on design recall distortion are used here.

It can be seen from Table 2 that both inability to complete the Word Learning Test in five trials and a design recall distortion score of 4 or more come near to such an ideal. Nevertheless the degree of certainty of differential diagnosis varies according to the alternative diagnosis to brain damage. Neither word learning nor design recall distortion, judging by the data here, are reliable diagnostic tests by themselves in certain psychiatric groups.

Table 2. *Percentage of subjects in each group obtaining various scores*

	O	AS	H	P	M	S	CS	EP
1. Word learning incomplete after five trials	81	18	30	10	23	33	70	40
2. Design recall distortion score more than 3	83	14	25	3	40	21	74	7
3. Design recall distortion score more than 6	25	0	0	0	13	4	13	0
4. Patients with both 1 and 2	66	4	10	0	10	8	57	7
5. MHVS-CPM discrepancy score more than +7	32	4	5	0	10	25	39	13
6. Patients with none of 1, 2 or 5	2	64	50	87	56	50	13	47
7. Patients with one of 1, 2 or 5	25	32	40	13	30	25	22	47
8. Patients with two of 1, 2 or 5	49	4	10	0	7	21	35	7
9. Patients with 1, 2 and 5	25	0	0	0	7	4	30	0

If a 'brain damage' score (as defined above) on both measures by one patient is made the criterion required, this largely removes the false positives. But only at the expense of increasing the number of organic cases falsely negative. Even so, if a patient gives 'brain damage' scores on both measures, he is almost certainly an organic case. The amount of overlap is sufficiently low as to be attributable to errors in psychiatric diagnosis.

A discrepancy score of greater than plus 7 as a measure of intellectual impairment was determined by an examination of the distributions summarized in Table B of the Appendix. It can be seen that this is less specifically a measure of brain damage than the other two measures. As relatively large proportions of both organics and schizophrenics show such a discrepancy score, it can best be considered a measure of intellectual impairment. It will also be noted from the table that the psychopaths are the only group to show a majority of minus discrepancy scores (i.e. non-verbal performance is better than verbal). Such a finding has tended to be reported more often than not with psychopathic and/or delinquent groups (see Faber, 1961).

Items 6, 7, 8 and 9 of Table 2 show the varying incidence of subjects giving critical scores in 0, 1, 2 or 3 measures (i.e., word learning incomplete on five trials, design recall distortion score more than 3, MHVS-CPM greater than +7). Again, different degrees of probability exist according to the group being compared with the organic group. Subjects with one score greater than the cut-off points occur with about the same incidence in all groups. For the occurrence of none or three scores beyond the cut-off points, however, there is virtually no overlap between the organic and any other group. It must be noted that the establishment of such a high level of

test certainty is only possible by having a 30-40 percentage of false negatives. This probably reflects the limitations of both the present techniques and of psychiatric diagnosis.

#### *Chronic schizophrenics and epileptics*

The chronic schizophrenics' performances on the Word Learning Test and with the design recall distortion score are not different from the organic group. From a practical point of view chronic schizophrenics are usually known to be such, but theoretically it does raise the question of the significance of such scores in chronic schizophrenia.

It is possible that these findings with chronic schizophrenia help to reconcile the apparently contradictory findings of Foulds & Dixon (1962), Foulds *et al.* (1962), Schwartzman & Douglas (1962) and Schwartzman *et al.* (1962). The former studies, using the Standard Progressive Matrices and the Mill Hill Vocabulary Scale, found no evidence that intellectual impairment in schizophrenia was progressive (after initial impact) with length of illness. The latter studies, however, (with different measures) did find such evidence. Perhaps indicators of intellectual impairment of a type non-specific for brain damage do not show any progressive increase in schizophrenia. Indicators more specific to brain damage might well suggest the presence of organic pathology in chronic schizophrenia (cf. Slater, 1963).

The twenty-three chronic schizophrenics studied here were all female with a range of hospitalization from 2-31 years (median 11 years). Out of a random total of eighty chronic schizophrenics interviewed, only these twenty-three performed all required tests adequately. Therefore, the present findings result only from the best chronic schizophrenics. The exclusion of the other fifty-seven patients was almost wholly due to doubts about the validity of their performances on the Word Learning Test and on design recall.

The epileptic group were psychiatric patients, suffering from idiopathic epilepsy, with various diagnostic labels other than that of organic. It is not really possible to say much about these findings here as larger groups are needed to control for psychiatric, epileptic and medication differences. It seems of interest, however, that it is only on the Word Learning Test that a high incidence of 'brain damage' performances occur. This is just possibly related to left-sided disturbances, perhaps temporal in origin. Various studies, in fact, have suggested a differential effect in the site of lesions according to varying results on verbal as opposed to visuo-spatial material (Meyer & Jones, 1957; McFie & Piercey, 1961; Doehring & Reitan, 1962).

#### *Positional learning effects*

The word learning test of ten words constitutes a task which might reveal positional and serial learning effects. There is considerable evidence that serial learning in normals results in the familiar bow-shaped curve, although explanations and expositions of this vary (Osgood, 1953; Jensen, 1962; Feigenbaum & Simon, 1962). With abnormals (Malmo & Amsel, 1948; Hall & Crookes, 1951, 1952), apart from a general finding that neurotics are worse overall than normals, no specific serial differences have been consistently reported.

Table 3 gives the incidence of success at any one of five points as a percentage of the total successes for any one group. The five points represent the ten items taken together in pairs so as to smooth out curve irregularities.

It can be seen that the psychopathic group are the only group producing anything like a

Table 3. *Positional effects on the Word Learning Test*  
Percentage success of total success for any group at 5 points

	1	2	3	4	5
Organics	35	19	17	19	10
Anxiety states	26	21	21	16	16
Hysterics	20	20	23	19	18
Psychopaths	27	22	16	15	20
Melancholics	20	20	23	20	17
Acute schizophrenics	33	23	17	14	13
Chronic schizophrenics	31	16	16	22	15



bow-shaped curve. Referring back to Table 2 it can also be seen that overall they are the most 'normal' scoring of any of the groups.

Although the other groups perform rather differently there is, in fact, no statistically significant difference between any of the groups as regards the shape of curve. An examination of Table 3 indicates that this is largely because points 2, 3 and 4 in the curve are very similar for all groups. It is the starting and finishing points that differ. In fact, the organics obtain a greater proportion of success at point 1 than either the hysterics or melancholics ( $\chi^2$ , d.f.i., Yates correction, = 5.78 and 5.67,  $p < 0.02$ ). This is true also of the schizophrenics ( $\chi^2 = 4.33$ , 4.80,  $p < 0.05$ ). At point 5 the organics obtain a smaller proportion of their success than psychopaths ( $\chi^2 = 5.55$ ,  $p < 0.02$ ). The differences between the groups, therefore, seems to be that organics, and to a lesser extent schizophrenics, fall off quite steadily in success throughout the positions on the Word Learning Test. Hysterics and melancholics, on the other hand, perform at the same level throughout. These four groups do not therefore produce a bow-shaped curve. These different effects might well be studied in future to examine their nature and significance.

#### *Design recall errors other than distortion*

Table 4 gives the incidence on the design recall test of errors other than distortion. Rotation, either in block design or design reproduction, is often considered a sign of organic pathology (e.g. Williams *et al.*, 1956). It has also been considered a more peculiarly epileptic phenomenon (Hovey, 1961). With the present material, however, it can only be considered a non-specific indicator of intellectual impairment.

Table 4. *Design recall errors other than distortion*  
(as percentage of each group)

	O	AS	H	P	M	S	CS	EP
Rotations	68	59	60	23	53	70	70	47
Omissions	30	32	10	17	26	25	39	13
Separations	32	32	55	4	23	29	91	33
Perseverations	34	7	0	7	4	14	13	0
Overlying	10	7	0	0	7	4	0	0

One noteworthy feature of Table 4 is the very high incidence of separation of parts of the design in the chronic schizophrenics. The only other high scoring group for separations is the hysteric one. Even so, the chronic schizophrenic incidence is significantly greater than the hysteric one ( $\chi^2$  6.03, d.f.1. Yates correction,  $p < 0.01$ ).

#### CONCLUSIONS

The test techniques used here are simple, take little time and are applicable to the great majority of subjects. The measures can usually be obtained from a patient in the course of one interview. The present results emphasize the usefulness of this in that, separately, none of the techniques provides a reliable diagnostic instrument of brain damage. Previous publications on the Word Learning Test and on design recall distortion separately, gave a much smaller degree of overlap between organics and non-organics. The difficulty appears to be that subjects in certain diagnostic categories are liable to score in what has previously been established as the brain-damage range. This feature has perhaps been obscured in the past by the practice of using mixed psychiatric groups as non-organic controls. Only by using the techniques in combination can the number of patients falsely positive for brain damage be reduced. This inevitably results in the non-identification of some brain-damaged patients by these tests. Therefore there are limitations to the usefulness of these

techniques. It might be fairly added, however, that these limitations probably include the more general limitations of psychiatric diagnosis. The concept of non-organic processes in the so-called functional disturbances, especially psychotic states, is difficult. This is particularly underlined by the present results in chronic schizophrenia, without attempting to avoid in any way the shortcomings of the psychological techniques used.

## REFERENCES

- DOEHRING, D. C. & REITAN, R. M. (1962). Concept attainment of human adults with lateralized cerebral lesions. *Percept. Mot. Skills*, **14**, 27-33.
- FEIGENBAUM, E. A. & SIMON, H. A. (1962). A theory of the serial position effect. *Brit. J. Psychol.* **53**, 307-320.
- FISHER, G. M. (1961). Discrepancy in verbal and performance I.Q. in adolescent sociopaths. *J. clin. Psychol.* **17**, 60.
- FOULDS, G. A. & DIXON, P. (1962). The nature of intellectual deficit in schizophrenia. I. A comparison of schizophrenics and neurotics. *Brit. J. soc. clin. Psychol.* **1**, 7-19.
- FOULDS, G. A., DIXON, P., MCCLELLAND, M. & MCCLELLAND, W. J. (1962). The nature of intellectual deficit in schizophrenia. II a cross-sectional study of paranoid, catatonic, hebephrenic and simple schizophrenics. *Brit. J. soc. clin. Psychol.* **1**, 141-149.
- FOULDS, G. A. & DIXON, P. (1962). The nature of intellectual deficit in schizophrenia. III A longitudinal study of the sub-groups. *Brit. J. soc. clin. Psychol.* **1**, 199-207.
- HALL, K. R. L. & CROOKES, T. G. (1951). Studies in learning impairment. I schizophrenics and organic patients. *J. ment. Sci.* **97**, 725-737.
- HALL, K. R. L. & CROOKES, T. G. (1952). Studies in learning impairment. II psychoneurotic patients. *J. ment. Sci.* **98**, 273-279.
- HOVEY, H. B. (1961). An analysis of figure rotations. *J. cons. Psychol.* **25**, 21-23.
- JENSEN, A. R. (1962). Is the serial position curve invariant? *Brit. J. Psychol.* **53**, 159-166.
- McFIE, J. & PIERCY, M. F. (1952). Intellectual impairment with localized cerebral lesions. *Brain*, **75**, 292-311.
- MALMO, R. B. & AMSEL, A. (1948). Anxiety produced interference in serial rote learning with observations on rote learning after partial pre-frontal lobectomy. *J. exp. Psychol.* **38**, 440-453.
- MEYER, V. & JONES, H. G. (1957). Patterns of cognitive test performance as functions of the lateral localization of cerebral abnormalities in the temporal lobe. *J. ment. Sci.* **103**, 758-772.
- ORME, J. E. (1957). Non-verbal and verbal performance in normal old age, senile dementia and elderly depression. *J. Geront.* **12**, 408-413.
- ORME, J. E. (1961). The Coloured Progressive Matrices as a measure of intellectual subnormality. *Brit. J. med. Psychol.* **34**, 291-292.
- ORME, J. E. (1962). Bender design recall and brain damage. *Dis. nerv. Syst.* **23**, 329-330.
- OSGOOD, C. E. (1953). *Method and Theory in Experimental Psychology*. New York: Oxford University Press.
- RAVEN, J. C. (1958). *Guide to the Coloured Progressive Matrices*. London: H. K. Lewis.
- RIDDELL, S. A. (1962). The relationships between tests of organic involvement, memory impairment and diagnosis in elderly psychiatric patients. *Brit. J. soc. clin. Psychol.* **1**, 228-231.
- SCHWARTZMAN, A. E. & DOUGLAS, V. I. (1962). Intellectual loss in schizophrenia. I. *Canad. J. Psychol.* **16**, 1-10.
- SCHWARTZMAN, A. E., DOUGLAS, V. I. & MUIR, W. R. (1962). Intellectual loss in schizophrenia. II. *Canad. J. Psychol.* **16**, 161-168.
- SLATER, E. & BEARD, A. W. (1963). The schizophrenia-like psychoses of epilepsy. V discussion and conclusions. *Brit. J. Psychiat.* **109**, 143-150.
- WALTON, D. (1958). The diagnostic and predictive accuracy of the Modified Word Learning Test in psychiatric patients over 65. *J. ment. Sci.* **104**, 1119-1122.
- WALTON, D. & BLACK, D. A. (1957). The validity of a psychological test of brain damage. *Brit. J. med. Psychol.* **30**, 270-279.



- WALTON, D. & BLACK, D. A. (1959). The predictive validity of a psychological test of brain damage. *J. ment. Sci.* **105**, 807-810.
- WALTON, D. & MATHER, M. D. (1961). A further study of the predictive validity of a psychological test of brain damage. *Brit. J. med. Psychol.* **34**, 73.
- WALTON, D., WHITE, J. G., BLACK, D. A. & YOUNG, A. J. (1959). The Modified Word Learning Test. A cross validation study. *Brit. J. med. Psychol.* **32**, 213-220.
- WILLIAMS, H. L., LUBIN, A., GIESEKING, C. & RUBINSTEIN, I. (1956). The relation of brain injury and visual perception to block design rotation. *J. cons. Psychol.* **20**, 275-280.
- YATES, A. J. (1954). The validity of some psychological tests of brain damage. *Psychol. Bull.* **51**, 359-379.
- YATES, A. J. (1956). The use of vocabulary in the measurement of intellectual deterioration, a review. *J. ment. Sci.* **102**, 409-440.

Manuscript received 4 April 1963

## APPENDIX

Table A. *MHVS—CPM expected discrepancies*

MHVS Percentile	Age				
	14	20-25	30-45	50-65	70-85
90	57(31)	73(37)	76(43)	77(45)	76(45)
90	53(27)	69(35)	73(40)	73(41)	71(40)
85	51(25)	67(32)	70(37)	71(39)	68(38)
80	48(22)	65(30)	68(35)	68(37)	66(36)
75	45(11)	63(29)	66(34)	65(35)	64(35)
70	44(10)	61(27)	64(32)	63(33)	62(34)
65	43( 9)	59(26)	62(31)	61(32)	60(32)
60	41( 8)	58(25)	60(29)	59(31)	58(31)
55	39( 7)	56(24)	59(28)	57(30)	56(30)
50	38( 6)	54(23)	58(27)	55(27)	54(28)
45	37( 5)	51(20)	55(25)	53(26)	50(26)
40	36( 5)	49(17)	53(23)	50(24)	47(25)
35	35( 4)	47(16)	51(22)	47(22)	44(22)
30	32( 3)	45(14)	49(20)	44(20)	41(21)
25	31( 3)	43(13)	47(19)	41(18)	38(20)
20	29( 2)	39(10)	42(15)	34(14)	32(18)
15	27( 2)	35( 7)	37(12)	29(11)	27(16)
10	25( 1)	30( 6)	32( 8)	25( 5)	21(14)
5	19( 1)	23( 1)	26( 4)	20( 3)	16( 6)

Figures in brackets are the expected discrepancies for each age and MHVS level.

Table B. *Percentage of subjects in different groups with various discrepancy scores*

	Discrepancy scores				
	Less than -7	-4 to -7	-3 to +3	+4 to +7	+8 and above
O	2	4	41	21	32
AS	3	14	57	23	3
H	0	5	65	25	5
P	10	33	47	10	0
M	0	20	40	30	10
S	0	4	56	16	24
CS	0	13	17	25	45
EP	0	13	61	13	13

## Neuroticism and Extraversion in the MMPI: Empirical Validation and Exploration\*

By NORMAN L. CORAH

*Dept. of Psychology, Washington University, St Louis 30, Missouri, U.S.A.*

The present study was an attempt to test the hypothesis that the two recurring MMPI factors are neuroticism (*N*) and introversion-extraversion (*I-E*). An initial factor analysis of several MMPI scales was carried out using college student *Ss*. Factor scores for the first two factors were obtained for seven groups of neurotics and sociopaths and two groups of normals. The scores on the first factor significantly differentiated neurotics from normals and sociopaths from normals. These scores did not differentiate the neurotic groups from each other.

The scores estimating the second factor were used in relation to Eysenck's hypothesis that patient groups differ in *I-E*. If the second factor dimension were *I-E*, these scores should differentiate dysthymic neurotics from hysterics and sociopaths, and normals should fall in the middle of the two extremes. This hypothesis was confirmed and it was concluded that the two factors were *N* and *I-E*. It was also found that when neurotic groups were selected on the basis of extreme scores on the second factor, their clinical diagnoses differed significantly in line with Eysenck's hypothesis.

An additional hypothesis that extraverted neurotics would be characterized more by somatic symptoms of anxiety while introverted neurotics are characterized by cognitive anxiety symptoms was tested by selecting two groups of neurotics with extreme *I-E* scores and comparing their reported symptoms. This hypothesis was also confirmed.

A number of factorial studies of the Minnesota Multiphasic Personality Inventory (MMPI, Hathaway & McKinley, 1951) have generally yielded two major factors which have been interpreted in various ways. Welsh (1956) noted a marked consistency in several of these studies and interpreted the dimensions in terms of a general factor of maladjustment or anxiety and a second factor of repression or denial. His Anxiety (*A*) and Repression (*R*) scales were developed from internal consistency criteria to measure these dimensions. Kassebaum, Couch & Slater (1959) included many non-clinical scales in their factor analysis of the MMPI and obtained two factors, a general factor of 'Neuroticism' or 'ego-weakness' and a second factor of introversion-extraversion. They noted that their results were consistent with Welsh's (1956) factor dimensions of *A* and *R*.

The assignment of various trait names to factors, usually on the basis of the items or tests which have the highest loadings on them, has always been one of the major problems in the use of factor analysis. Whenever it is possible, the investigator should go one step further and attempt to demonstrate that the traits presumed to underly the factors are, in fact, what is being measured. There should be little problem with a measure presumed to represent the trait of neuroticism or ego-weakness since it should differentiate normal *Ss* from those diagnosed as neurotics.

However, in the case of a factor reputed to be introversion-extraversion, no clearly acceptable external criterion exists for the trait dimension. The only alternative in

\* An earlier version of this paper was read at the Midwestern Psychological Association Meeting in Chicago, May 1963.



this instance is to make an attempt to establish the construct validity of the dimension. Eysenck (1960c) has argued, on the basis of Jungian theory, that dysthymic neurotics (obsessives, depressives, anxiety neurotics) are introverted while hysterics and psychopaths are extraverted neurotics. Normal Ss are presumed to fall in the middle of these two extremes. Consequently, a purported measure of introversion-extraversion should significantly differentiate dysthymic neurotics from normals, hysterics and psychopaths but should not differentiate among dysthymic groups. Normals should also be significantly differentiated from hysterics and psychopaths. The present study attempts to examine the factor dimensions within this framework.

## METHOD

### *The factor measures*

It was first necessary to obtain factor score weights. An attempt was made to replicate the Kassebaum *et al.* (1959) study using only a few MMPI scales which were known to have high loadings on these factors: *Pt*—Psychasthenia; *D*—Depression; *K*—Ego functioning; *A*—Welsh's anxiety scale (1956); *Es*—Barron's ego strength scale (1953); *R*—Welsh's scale of repression and denial (1956); *Ex*—Giedt and Downing's extraversion scale (1961) which was derived from *Ma* (Hypomania) and Gough's sociability scale (1957).

Two additional scales that presumably deal with defensiveness were used: *L*—Lie, and *HyDn*—Little and Fisher's hysteria denial scale (1958).

A 250 item short form MMPI including these nine scales was administered to 180 college students, eighty-four men and ninety-six women, from general introductory psychology classes at Washington University. The scales were intercorrelated and the matrix factor analysed by the centroid method. Three factors, accounting for 69 per cent of the total variance, were extracted and graphically rotated to orthogonal simple structure. The rotated factor matrix is presented in Table 1. The first two factors are clearly identical with those of Kassebaum *et al.* (1959).

Table 1. *Rotated factor loadings*

Test	I	II	III	$h^2$
<i>Pt</i>	0.92	0.00	-0.12	0.86
<i>D</i>	0.58	0.56	0.06	0.65
<i>HyDn</i>	-0.43	0.06	0.84	0.89
<i>L</i>	-0.13	0.33	0.53	0.41
<i>K</i>	-0.61	0.30	0.52	0.73
<i>A</i>	0.93	0.01	-0.20	0.90
<i>R</i>	-0.14	0.79	0.25	0.71
<i>Ex</i>	-0.19	-0.76	0.15	0.64
<i>Es</i>	-0.65	-0.13	0.00	0.44
$\Sigma a^2 =$	3.10	1.73	1.40	$\Sigma h^2 = 6.23$

In order to obtain factor scores, it was first necessary to obtain the multiple regression coefficients. Several combinations were worked out for each factor using Aitken's method of pivotal condensation (Thompson, 1950). No combination of tests for the first factor added any significant weight to that which already obtained for either *A* or *Pt* alone. The distribution of *A* scores was somewhat more skewed than that of *Pt*. Consequently, *Pt* was used as the best measure of Factor I. The best combination for the second factor was  $0.5z_R - 0.5z_{Ex}$  which gave an estimated multiple correlation of 0.88 with Factor II.

### *Test of the hypothesis*

MMPI records were obtained for 251 Ss from nine different groups: two 'normal' groups, one of thirty-eight college students from Washington University and one of seventy-two psychiatric aides and attendants; five out-patient neurotic groups consisting of nine phobic

reactions, twenty obsessive-compulsives, sixteen depressives, twenty-two hysterics and thirty-nine anxiety neurotics; two in-patient groups with a primary diagnosis of sociopathic personality consisting of twenty-four patients with a sub-classification of anti-social reaction and eleven with no sub-classification ('sociopaths'). The records of the second normal group and all of the patient groups were obtained from St Louis State Hospital.\* Classifications were made on the basis of hospital staff diagnoses.

The factor scores were obtained for all cases and converted to T-scores. The standard conversion table for *Pt* (Hathaway & McKinley, 1951, p. 16) was used for the first set of scores. A special table of T-scores was used for Factor II and was based on the data for the original group of 180 college students. The distributions were separate for each sex since college women had obtained significantly higher factor scores ( $t = 3.00$ , d.f. = 178,  $p < 0.005$ ).

Age was an uncontrolled variable since there was a significant difference among the groups in age ( $F = 5.65$ , d.f. = 7/213,  $p < 0.001$ ) with the college students eliminated from the analysis. However, age did not correlate significantly with the first factor scores with an  $r$  of  $-0.04$ . The correlation between Factor II and age was  $0.18$  which was significant ( $p < 0.01$ ). The Factor II measures were corrected by partialing out the effects of age. The two sets of factor scores were correlated to test for independence. An  $r$  of  $0.22$  was obtained which is significant ( $p < 0.01$ ). Consequently, an analysis of covariance was used in assessing each factor independently.

## RESULTS

The covariance analysis of the factor scores is presented in Table 2. Both analyses are significant beyond the  $0.001$  level. The adjusted group means for both factor dimensions are presented in Fig. 1.

Table 2. *Analysis of covariance for the factor scores among the eleven groups with the influence of each factor partialled out of the analysis of the other*

Source	d.f.	ss <sub>I</sub>	sp	ss <sub>II</sub>	d.f.	ms <sub>I</sub>	ms <sub>II</sub>	$F_I$	$F_{II}$
Group	8	23,371.02	8,097.56	4,731.13	8	2,604.03	432.32	20.22*	4.62*
W	242	31,032.94	228.55	22,550.46	241	128.76	93.56		
Total	250	54,403.96	8,326.11	27,281.59	249				

\* $p < 0.001$ .

The two 'normal' groups differed significantly on the first factor scores ( $p < 0.01$ ) but not on the second set of scores. However, since the college students were above the mean of the norm group for *Pt* and the psychiatric aides were below it, the two groups were combined for the  $t$ -tests between groups. All of the patient groups except the 'sociopaths' differed significantly from the normals ( $p = 0.05$  or less) on Factor I. None of the neurotic groups differed significantly from each other but differed significantly ( $p = 0.05$ ) from the sociopathic groups. There was no significant difference on this factor between the sociopathic groups. This first factor would seem to be reasonably well identified with the trait of neuroticism ( $N$ ).

The scores on Factor II significantly ( $p < 0.05$ ) differentiated the dysthymic groups (obsessives, depressives, anxiety neurotics and phobics) from the hysterics, the dysthymics from the normals ( $p < 0.01$ ), the normals from the sociopaths ( $p < 0.01$ ), and the hysterics from the sociopaths ( $p < 0.05$ ). None of the differences among the dysthymic groups or between the sociopathic groups was significant. The results

\* The writer wishes to thank Dr Abel Ossorio, Director of the Division of Psychology and Dr Robert Scheaf, Chief Psychologist of the Outpatient Service at St Louis State Hospital for their aid in making records available.



## NEUROTICISM

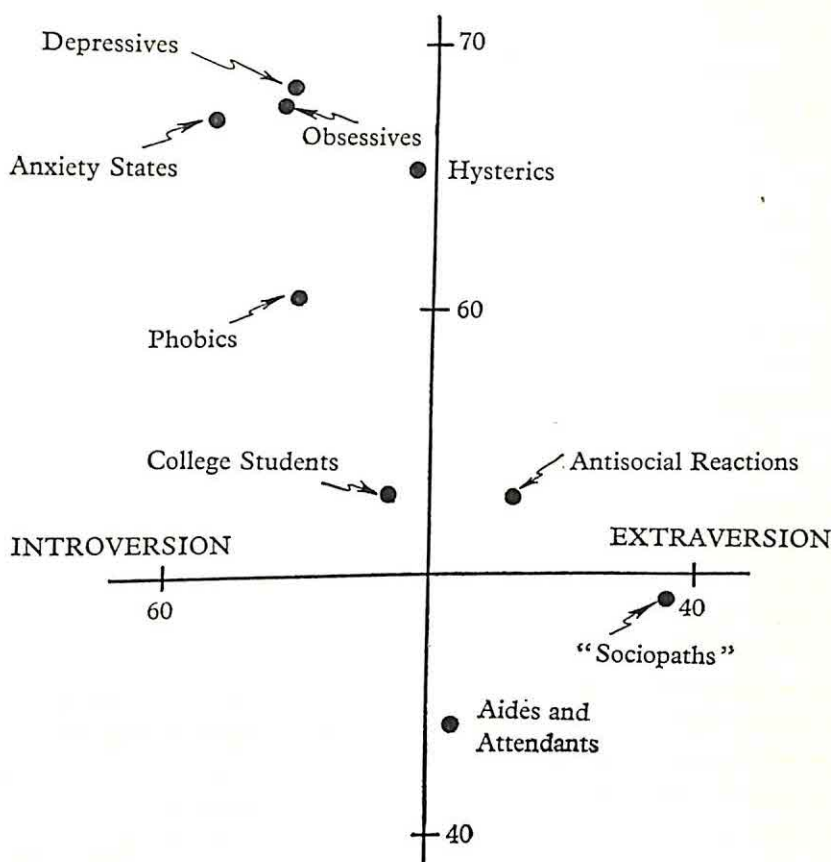


Fig. 1. Adjusted means of the nine groups for both factor dimensions.

with the scores for the second factor appear, for the most part, to meet Eysenck's (1960c) criterion for a bipolar trait of introversion-extraversion (*I-E*).

#### *An extension and further validation*

Eysenck (1960b), in reviewing some studies of symptom ratings, has hypothesized that extraverted neurotics will be characterized primarily by somatic manifestations of anxiety while introverted neurotics will be characterized by 'psychic' or cognitive manifestations of anxiety. The next study is a test of this hypothesis and, in addition, represents an attempt to provide further validation for the *I-E* dimension.

#### METHOD AND RESULTS

From a group of 128 out-patient neurotics (including twelve sociopathic diagnoses) at St Louis State Hospital, twenty-five extreme extraverts (ten men and fifteen women) and twenty-five extreme introverts (eleven men and fourteen women) were selected on the basis of their *I-E* factor scores. The mean *N* factor scores of the groups were 68.88 for the extraverts and 71.32 for the introverts. The difference was not significant. The mean *I-E* scores were 41.84 for the extraverts and 70.72 for the introverts.

An associate of the writer, who did not know the purpose of the study, read the accounts of the social history and psychiatric interview of each of the fifty patients and noted each anxiety symptom reported by the patients.\* The lists of symptoms were then rated by three judges independently, without knowledge of the patient's group, as to whether each symptom indicated somatic or cognitive anxiety. Somatic symptoms included autonomic complaints, somatic complaints, motor activity and acting out. Cognitive symptoms included feelings of tension, fears, depression, anxious mood, insomnia, 'bad' dreams and nervousness.

The three raters were in complete agreement on 91 per cent of the total of 409 symptoms. On symptoms where there was disagreement, the symptom was assigned to the category on which two of the three raters agreed. A 'proportion of somatic anxiety' score was obtained for each patient by subtracting the number of cognitive symptoms from the number of somatic symptoms and dividing by the total number of symptoms. The mean scores were 0.5152 for the extraverts and 0.2809 for the introverts. The difference is significant beyond the 0.001 level ( $t=4.105$ , d.f. = 48).

Further support for the contention that the second factor represents *I-E* might be obtained by testing the hypothesis that the diagnoses of the two groups of patients would fall into the respective categories of dysthymic neurotics and extraverted neurotics as Eysenck has postulated (1960c). There were, however, ten patients with the diagnosis of Passive-Aggressive Personality in these groups, and it was not exactly clear as to which general neurotic category they belonged. It was arbitrarily decided to classify these patients on the basis of the sub-diagnosis of 'Passive type' as introverted neurotics and 'Aggressive type' as extraverted neurotics.

The diagnostic classifications of the patients are presented in Table 3. The diagnoses were combined according to Eysenck's criteria into two groups—introverted neurotics and extraverted neurotics. This resulted in a two by two contingency table. A chi square of 11.68 (d.f. = 1,  $p < 0.001$ ) was obtained which was significant.

Table 3. *Diagnostic classification of the fifty cases selected on the basis of their I-E factor scores*

Diagnosis	Number of cases in each of the extreme factor score groups	
	Introverts	Extraverts
Anxiety Reaction	8	4
Obsessive-Compulsive	1	1
Depressive	3	1
Phobic	1	1
Passive-Aggressive, passive	7	1
Passive-Aggressive, aggressive	1	1
Hysteric	3	10
Sociopath	1	6
Total	25	25

\* The writer wishes to thank Mrs Deborah S. Onken for her assistance in the collection and analysis of part of the data.



## DISCUSSION

The results of this research appear to substantiate the assumption that the MMPI factor dimensions studied are *N* and *I-E*. There are, however, two aspects of the results which should be mentioned. The sociopathic groups tend to be closer to the normals on the *N* dimension than to the neurotics. This result is not consistent with Eysenck's theorizing. However, it is consistent with standard nosological criteria (American Psychiatric Association, 1952) in which sociopathic personality is considered separate from the neuroses and psychoses. The second result which is inconsistent with Eysenck's theory is that in which the hysterics do not differ from the normals in *I-E*. This result is consistent with other data and Eysenck (1959, p. 6) has noted that this finding '... has been repeated on several samples and must be accepted'. It might be added that the predicted difference between hysterics and normals has always taken second place to the postulated difference between dysthymics and hysterics.

The finding that patients selected on the basis of the second factor dimension significantly differ in their clinical diagnoses, in line with the theoretical classification of extraverted and introverted neurotics, lends additional support to the contention that the second factor is an *I-E* dimension.

Eysenck (1962) has suggested that *Pt*, the measure of *N* used in the present study, is a mixture of *N* with introversion. However, it may be noted that *Pt* has had a higher loading on a neuroticism factor than his own *N* scale in at least two factor analytic studies (Eysenck, 1960a, 1962).

The old criticism of applying results obtained from college students to another population might be raised in connection with the present study. It was felt that the results would be more impressive if the initial factor scores were derived from a different group. However, the writer has since replicated the factor analysis of the nine scales on a group of 154 neurotic patients. *Pt* had the highest loading (0.97) on Factor I. The weights obtained from the college students for the second factor were applied to this group. The resulting estimated multiple correlation of *R* and *Ex* with Factor II was 0.86—almost identical to that obtained for the students.

The second part of the present study clearly substantiates Eysenck's (1960b) hypothesis that extraverted neurotics will show a preponderance of somatic symptoms of anxiety while introverted neurotics will be characterized more by cognitive symptoms of anxiety. Such findings lend emphasis to Eysenck's (1960c) contention that *both* dimensions (*N* and *I-E*) must be taken into account in studies of neuroticism.

## REFERENCES

- AMERICAN PSYCHIATRIC ASSOCIATION, Mental Hospital Service, Committee on Nomenclature and Statistics of the American Psychiatric Association. (1952). Diagnostic and statistical manual: *Mental disorders*. Washington, D.C., APA.
- BARRON, F. (1953). An ego-strength scale which predicts response to psychotherapy. *J. cons. Psychol.* **17**, 327-333.
- EYSENCK, H. J. (1959). *Manual of the Maudsley Personality Inventory*. London: University of London Press.
- EYSENCK, H. J. (1960a). Classification and the problems of diagnosis. In *Handbook of Abnormal Psychology* (H. J. Eysenck, ed.), pp. 1-31. New York: Basic Books.

- EYSENCK, H. J. (1960b). *Experiments in Personality*, Vol. 2. London: Routledge & Kegan Paul.
- EYSENCK, H. J. (1960c). *The Structure of Human Personality*, 2nd ed. London: Methuen.
- EYSENCK, H. J. (1962). Response set, authoritarianism and personality questionnaires. *Brit. J. soc. clin. Psychol.* **1**, 20-24.
- GIEDT, F. H. & DOWNING, L. (1961). An extraversion scale for the MMPI. *J. clin. Psychol.* **17**, 156-159.
- GOUGH, H. G. (1957). *Manual for the California Psychological Inventory*. Palo Alto, California: Consulting Psych. Press.
- HATHAWAY, S. R. & MCKINLEY, J. C. (1951). *Minnesota Multiphasic Personality Inventory Manual*. New York: Psychological Corp.
- KASSEBAUM, G. G., COUCH, A. S. & SLATER, P. E. (1959). The factorial dimensions of the MMPI. *J. cons. Psychol.* **23**, 226-236.
- LITTLE, K. B. & FISHER, J. (1958). Two new experimental scales of the MMPI. *J. cons. Psychol.* **22**, 305-306.
- THOMPSON, G. H. (1950). *The Factorial Analysis of Human Ability*, 4th ed. New York: Houghton Mifflin.
- WELSH, G. S. (1956). Factor dimensions A and R. In *Basic Readings on the MMPI in Psychology and Medicine* (G. S. Welsh & W. G. Dahlstrom, eds.), pp. 264-281. Minneapolis: University of Minnesota Press.

*Manuscript received 4 July 1963*



## A Study of 500 Oxford Student Psychiatric Patients

BY M. A. DAVIDSON AND CORINNE HUTT

*The Warneford Hospital, Oxford*

### SUMMARY

The case records and psychometric results of 500 Oxford student patients were retrospectively analysed. Some academic, social and clinical characteristics of the patient sample are described. The sample is also compared with a control group and with the student population at risk.

The sample was found not to differ significantly from other psychiatric groups. It is pointed out that a high referral rate need not be indicative of a serious problem of student 'breakdown'. There is no evidence to show that there are factors in the University environment which generate psychiatric illness.

Much concern and interest has been shown in the subject of psychological ill-health amongst students. Parnell (1951) reported that mental illness had caused more than half the losses of at least a term's residence of Oxford undergraduates. Wolfe (1954) reflected an international concern when he reported that a third of the brightest 20 per cent of the American male potential failed to achieve graduate status on grounds of psychological illness. Parnell (1953) reported that the admission rate to the Warneford Hospital, Oxford, was more than twice the national admission rate to psychiatric hospitals for men of the same age group and social class. Davidson *et al.* (1955) stated that 'the evidence as a whole shows that, as found in 1951, mental upset is still the most important cause of ill-health in Oxford students'.

This report is based on the findings of a retrospective study of 500 student psychiatric patients; its limitations are those of any retrospective analysis, where the data have been abstracted from case-notes written by several different clinicians. Nevertheless, a number of interesting findings have emerged.

### *The sample*

The patients were 500 male undergraduate and post-graduate students at Oxford University who had been admitted as in-patients, or had attended as out-patients at The Warneford Hospital, Oxford, and had been subsequently interviewed and psychometrically tested by the first author, over the period 1950-61. There were 198 in-patients and 302 out-patients, most of them referred by their general practitioner. Twenty-five were admitted after attempted suicide. The age range of the patients was 18-30 years inclusive, the mode of the distribution being at 21. Post-graduate students over the age of 30 were not included in the sample.

### *The data*

All patients' case-files were analysed in detail and information extracted under two main headings: (1) academic; and (2) social and clinical. All data were recorded without knowledge of the diagnosis, this being entered as the final item. The data were coded and put on to 85-column punched cards which were then sorted by an I.C.T. machine.

## RESULTS

The results are given in three sections:

- (1) Descriptive characteristics of the sample.
- (2) Comparisons with the student population at risk, and with a control group.
- (3) Analysis of the patient group.

Where information was not available on a particular variable for any student, he was omitted from the analysis of that variable. Thus N will not be identical in all computations.

1. *Descriptive characteristics**Academic*

(a) *Student status.* 439 (88 per cent) of the patients were undergraduates, 61 (12 per cent) post-graduates. 280 (56 per cent) were Commoners and of these 51 were State Scholars; 95 (19 per cent) were College Scholars and 37 (7.4 per cent) were Exhibitioners. Only three were both College and State Scholars.

(b) *Subject.* The distribution among the subjects read was as in Table 1.

Table 1. *Subjects read*

Arts			Science		
	N	%		N	%
Languages	39	7.8	Natural Science	58	11.6
Literae Humaniores	67	13.4	Medicine	25	5.0
Jurisprudence	27	5.4	Mathematics	12	2.4
Mod. History	58	11.6	Applied Sciences	13	2.6
Theology	16	3.2	P.P.P.	5	1.0
English	75	15.1			
P.P.E. (and Philosophy)	68	13.6		113	22.6
Music	12	2.4			
Geography	10	2.0	Diploma in Education	15	3.0
	372	74.4			

(c) *Examinations.* Of those who took their First Public examination, 31 per cent needed extra time to complete. Of those who passed their final examination 30 per cent needed more than the usual time permitted (4 years for Greats and Chemistry and 3 years for all other subjects) to take the examination.

(d) *Secondary schools.* 202 patients (40 per cent) had attended grammar school, 206 (41 per cent) had attended public school (i.e. members of Head-Masters Conference), 52 (10 per cent) had received their schooling abroad and the rest had attended some other type of school, e.g. independent or Scottish academy.

*Social*

(a) *Social class.* Using the Registrar General's categories of social class the distribution, according to parents' occupation, was:

I	II	III	IV & V
219 (44%)	185 (37%)	76 (15%)	17 (4%)

(b) *Social adjustment.* 130 patients (26 per cent) claimed that they were socially at ease, whereas 73 (14.6 per cent) stated they were uncomfortable, ill at ease socially in the University environment; 20 found their lowly family status an embarrassment and 28 were in financial difficulties.

*Clinical*

(a) *Symptoms and diagnosis.* The symptom(s) the patient complained of at the initial interview was recorded as the presenting symptom(s). The most common were depression



(44 per cent), work difficulty (42 per cent), and sexual problems (21 per cent—homosexual 14.6 per cent). By diagnosis the patients were classified as follows:

Schizophrenic psychoses (SCH): 43 (8.6 per cent)—schizophrenia, schizophrenic episode/prodrome.

Affective psychoses (AFF): 79 (15.8 per cent)—endogenous depression, hypomania, cyclothymia.

Anxiety states (ANX): 93 (18.6 per cent)—anxiety reactions, phobic or obsessional anxieties.

Neurotic depressions (ND): 104 (20.8 per cent)—reactive depressions.

Psychopathy and hysteria (P & Hy): 26 (5.2 per cent).

Personality disorders (PD) (other): 67 (13.4 per cent)—obsessional, immature, etc.

Sexual difficulties (SX): 60 (12 per cent)—largely homosexual problems.

Organic: 8 (1.6 per cent).

Twenty (4 per cent) were stated to have 'no formal psychiatric illness'.

(b) *Family psychiatric illness.* 100 patients (20 per cent) had parents who had been psychiatrically ill. 110 had relatives (both direct and collateral) who had been similarly ill.

(c) *Previous psychiatric disturbances.* 144 patients had manifested symptoms of disorder in childhood, e.g. enuresis, stealing, temper tantrums; 52 attended a child guidance clinic or had some other psychological investigation. 36 patients had already had psychiatric consultations, prior to attending this Hospital, and since commencing secondary school (two of these fell in the former group as well).

(d) *Time of first consultation.* Table 2 shows the time of first attendance for psychiatric consultation. The first attendance rates of the first 3 years (the majority of students do a 3-year course) were fairly similar; over the three terms of each year, however, the picture was different, there being a peak in each spring term. The null hypothesis that there would be no differences in the proportions attending each term was rejected at almost the 0.1 per cent level ( $\chi^2 = 13.4$ ).

Table 2. *Time of first consultation*

	1st year		2nd year		3rd year		4th year		5th year		Total	
	N	%	N	%	N	%	N	%	N	%	N	%
Michaelmas Term	50	10	39	8	46	9	16	3	1		152	30
Hilary Term	62	13	63	13	64	13	9	2	1		199	41
Trinity Term	54	11	33	7	38	8	9	2	1		135	28
Total	166	34	135	28	148	30	34	7	3	1	486	

## 2. Comparisons with the student population at risk, and with a control group

The control group consisted of 100 freshmen volunteers from three Colleges (Davidson *et al.*, 1955). Since nearly two-thirds of this group were from one College it was not truly representative, in many respects, of the student population. In addition it was also likely to be subject to the biases found in volunteer groups (Bell, 1962; Howe, 1960; Martin & Marcuse, 1958). Wherever possible, therefore, comparisons have been made with the relevant figures for the male student population at risk at Oxford University over the academic years 1950-51 to 1960-61. Table 3 compares the patient sample with the population at risk on three variables; student status, subject read and degree results. Except for the first two comparisons, only the undergraduate figures have been used. (The subjects listed are those which showed the greatest discrepancies between patient and population proportions.)

The patients did as well as the general undergraduate population in terms of 1st class degrees gained, but obtained fewer 2nds, and more 3rds and 4ths. The most able students, therefore, did not appear to be materially handicapped by their illness.

Undergraduates reading English were significantly over-represented in the patient sample and so, to a less extent, were those reading Music and Greats. Potential lawyers, however, contributed significantly few to the patient group. College scholars, too, were over-represented among the patients.

Table 3. *Comparisons between the patient sample and the male undergraduate population at risk*

Criterion	Pop. %	Patient	Diff. in props S.e. of diff.	Significance
1. <i>Student status</i>				
Undergraduates	82.6	88.2	3.06	**
Post-graduates	17.4	11.8		
Scholars	15.6	19.0	2.06	**
Exhibitioners	7.8	7.4		
2. <i>Subject</i>				
English	8.4	15.2	5.10	***
Music	0.7	2.1	3.20	**
Greats	10.5	15.0	3.06	**
Jurisprudence	11.2	6.0	3.40	***
Modern History	15.6	12.4	1.83	n.s.
Arts	73.8	77.6	1.81	n.s.
Science	26.2	22.4		
3. <i>Degree results</i>				
I	8.4	8.7		
II	52.3	46.0	2.20	*
III	34.1	37.3		
IV	5.2	8.0	2.10	*

\*\*\*  $p \leq 0.001$ ; \*\*  $p \leq 0.01$ ; \*  $p \leq 0.05$ .

Table 4. *Comparisons between the Patient sample and a Control Group*

Criterion	Patients		Controls		$\chi^2$	Sig.	C.C.
	N	%	N	%			
1. Poor relationships with parents:							
With mother	207	47	7	7	20.010	***	0.20
With father	248	55	11	11		49.800	***
2. Pattern of guardianship:							
Complete	404	81	87	87	1.760	n.s.	
Incomplete	96	19	13	13			
3. Parental disharmony							
	83	28	16	22	0.750	n.s.	
4. Home atmosphere interrupted							
	75	18	16	17	0.004	n.s.	
5. Sexuality:							
Masturbation	221	71	68	68	0.300	n.s.	0.12
Homosexual interests	144	38	21	23	7.140	**	
No heterosexual interests	37	9	7	7	0.040	n.s.	
6. Family psychiatric illness:							
Any relative	174	45	19	19	21.830	***	0.21
Direct (and collateral)	125	34	12	12	17.050	***	0.19
Collateral only	32	8	7	7		n.s.	
7. Previous psychological investigation							
	86	20	2	2	17.410	***	0.20
8. Secondary School:							
Public School (H.M.C.)	206	41	36	36	3.000	n.s.	
Grammar School	202	40	52	52			
Other school	92	19	12	12			

In all cases  $\chi^2$  for 1 degree of freedom and corrected for continuity (Yates).



Table 4 compares patient and control groups on eight main variables for which information was available from both groups. Poor relationships with both parents (those characterized by hostility, resentment or extreme ambivalence) were reported significantly more often by the patients than the controls, though the degree of association with a poor paternal relationship was considerably higher than that with a poor maternal relationship.

Patients and controls were classified as having a 'complete' pattern of guardianship if they had both parents alive and responsible for their rearing until the age of 12 years; an 'incomplete' pattern included those who had, for some reason or other, lost one or both parents, been adopted or fostered.

It will be seen that neither parental disharmony (i.e. parents quarrelling frequently, leading independent lives, or separated) nor the nature of the pattern of guardianship were factors differentiating significantly between patients and controls. These findings in general confirm the results of Davidson *et al.* (1955).

The home atmosphere was said to be interrupted when it had periodically altered, due to the fact that the father was employed in the service, foreign office, or some other such peripatetic post. The results show that interrupted home conditions were not significantly more frequent amongst the patients.

Masturbation and lack of heterosexual interests (i.e. no active interest in opposite sex) were not more common among the patients, but homosexual interests (friendships and/or active relationships) were. A morbid family history was significantly associated with the patient group, this association being almost entirely contributed to by direct relatives, i.e. a grand-parent, parent or sibling. Many more patients than controls had had previous psychological investigation.

There was no difference between the proportions in the two groups, that had attended grammar school and public school. The proportions of patients and controls attending boarding school were almost identical.

Table 5. *Test results: Test means and standard deviations*

Test	Patients (N = 392)		(N = 100)		<i>t</i>
	Mean	s.d.	Mean	s.d.	
<i>General knowledge</i>					
Professions test—Pt. I	43.4	11.02	41.4	11.30	n.s.
Pt. II	37.0	8.00	37.5	8.60	n.s.
Total	80.3	17.81	78.8	18.88	n.s.
<i>Intelligence</i>					
AH5 test—Pt. I	18.8	4.80	19.1	4.40	n.s.
Pt. II	18.6	5.80	20.2	5.00	3.03**
Total	37.4	9.21	39.2	8.31	2.07*

Table 5 shows the means and standard deviations of patients and controls on a test of general knowledge (Anstey, 1948) and on AH5, a test of high-grade intelligence standardized on a University population, (Heim, 1947). The two groups were well matched as regards general knowledge, but patients did less well than the controls on the intelligence test. The scores of the two groups on AH5, Part I, (verbal and arithmetical reasoning) were very little different, but the discrepancy on Part II (largely visuo-spatial and non-verbal) was a significant one, and in accordance with what might be expected from a psychiatrically disabled group (see, e.g., Payne, 1960).

Mean scores on the scales of the MMPI were calculated for the two groups. The composite MMPI profiles thus obtained are shown in Figure 1. There is a remarkable similarity between the two profiles from the Hysteria (Hy) scale onwards. The discrepancy between the mean scores on the Depression scale is significant at the 1 per cent level. The patient peak on Depression reflects the large proportion of diagnoses of depression, as well as the most common presenting symptom.

3. Analysis of the patient group

(1) Public school and grammar school patients

Although the distinction between public school and grammar school is far from clear, it is one often made in practice. Comparisons between these two groups of patients showed no significant differences as regards intelligence, academic achievements or homosexual

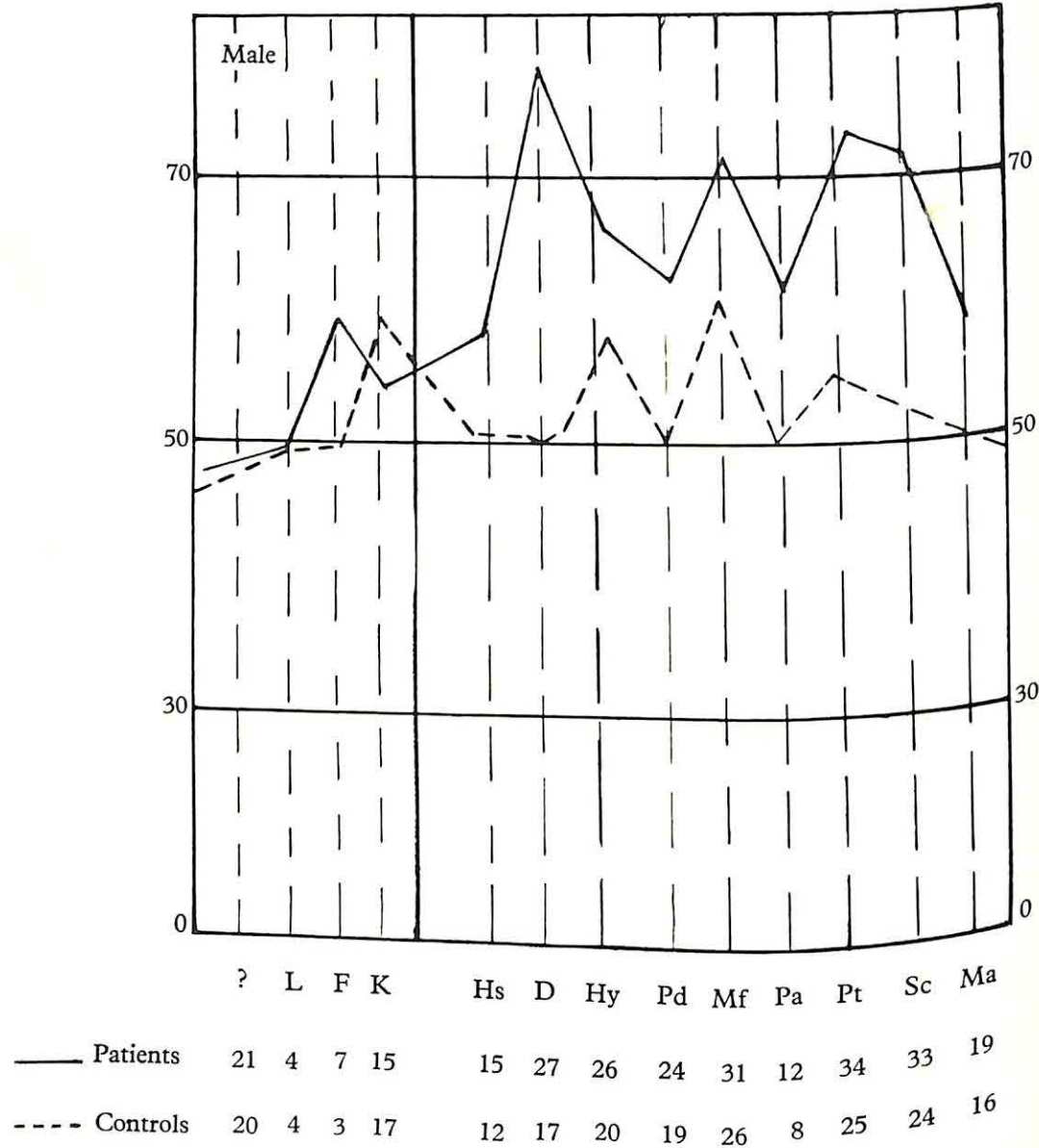


Fig. 1. Composite MMPI profiles for patients and controls.

interests. Significantly more public school patients, however, claimed to be socially at ease in the University environment. Homosexual interests were not found significantly more frequently among boarders than day-boys.

(2) Diagnostic groups

The different diagnostic groups (excluding the eight organic patients) were compared on four variables. In all cases where multiple comparisons of proportions were made, Ryan's (1960) test was used.



(a) *Childhood disorder.* Table 6 shows the proportions of patients in each diagnostic group who had manifested symptoms of psychiatric disorder in childhood. On the whole the psychotic and personality-disorder groups showed a significantly higher proportion of patients who had shown such symptoms, than the anxiety and reactive-depression groups.

Table 6. *Comparisons between proportions showing childhood symptoms*

Proportions	Comparisons	Significance
	P & Hy $\times$ ND	*
P & Hy: 0.44	P & Hy $\times$ ANX	*
SCH: 0.42	SCH $\times$ ND	*
PD: 0.39	P & Hy $\times$ SX	n.s.
AFF: 0.34	SCH $\times$ ANX	*
SX: 0.28	PD $\times$ ND	*
ANX: 0.23	P & Hy $\times$ AFF	n.s.
ND: 0.18	PD $\times$ ANX	n.s.
	AFF $\times$ ND	n.s.

(b) *Outcome—clinical.* Clinical outcome was recorded as (a) good when patient was stated by the clinician to have recovered, (b) fair when prognosis was uncertain, and (c) poor when patient was readmitted or sought psychiatric help again within the period of the study. The association between outcome and diagnosis was a highly significant one, the trend being in the direction to be expected of a psychiatric sample: the schizophrenics had a relatively unfavourable outcome, as did the personality-disorder groups, whereas the reactive states and SX group showed a good recovery (Table 7).

Table 7. *Clinical outcome by diagnosis*

Outcome	SCH	AFF	P & Hy	ANX	ND	PD	SX
Good	7	28	2	56	52	18	29
Fair	13	24	10	24	30	31	21
Poor	20	23	12	7	8	15	7

$$\chi^2 = 78.3***; 12 \text{ d.f.}; \text{c.c.} = 0.39.$$

(c) *Outcome—academic.* Table 8 shows the distribution of degrees in the various diagnostic groups. The schizophrenic psychotics frequently failed to get a degree and only infrequently obtained a good degree; the neurotic groups, however, did much better than would have been expected on the null hypothesis, more often than not obtaining a good degree.

Table 8. *Academic outcome by diagnosis*

	SCH	AFF	P & Hy	ANX	ND	PD	SX
1st and 2nd class degrees	9	25	6	40	44	18	19
Other degrees	16	19	11	29	28	31	25
No degree	15	13	6	7	6	8	3

$$\chi^2 = 41.83***; 12 \text{ d.f.}$$

(d) *Family psychiatric illness.* Table 9 indicates the proportions of patients in the various groups who had a morbid family history (direct relatives only) and their significant differences. The psychotic groups had a significantly greater proportion of family illness than the neurotic groups, with the personality-disorder groups falling in between.

### (3) *Period of symptom versus outcome*

A considerable number of patients (or their parents) provided the information that their symptoms had been evident for a period prior to their attendance at the hospital. When these symptom periods were considered in relation to the subsequent outcome, a very

Table 9. *Comparisons between proportions with morbid family history*

Proportions		Comparisons	Significance
SCH	0.51	SCH $\times$ SX	**
AFF	0.46	SCH $\times$ ND	*
PD	0.37	AFF $\times$ SX	*
P & Hy	0.33	SCH $\times$ ANX	*
ANX	0.27	AFF $\times$ ND	*
ND	0.24	PD $\times$ SX	n.s.
SX	0.19	SCH $\times$ P & Hy	n.s.

significant association was apparent; those patients whose symptoms had been evident for less than a year had a far better prognosis than the others; the longer the symptom had been in evidence the more resistant it appeared to be to treatment, as shown in Table 10.

Table 10

Outcome/Period	Less than 1 year	1-2 years	3-4 years	5 or more years
Good	35	27	32	18
Fair	16	13	29	34
Poor	1	9	17	21

$$\chi^2 = 30.27 \text{ ***; } 6 \text{ d.f.; C.C.} = 0.35.$$

#### (4) *Patients with work difficulty*

Since a large proportion of patients presented with the symptom 'work difficulty', it was decided to investigate this group further and to compare it with a group of patients who stated that they could cope adequately with their work. There was no evidence that work difficulty could be attributed to a difference in intelligence between the two groups as when their AH5 results were examined the 'work difficulty' group was seen to have the significantly higher scores (Table 11). This may, however, be due to the fact that the 'work difficulty' group consisted largely of patients with neurotic disorders, and were therefore less handicapped on the test than, say, the psychotics.

Table 11

	Work difficulty	No work difficulty
AH5 Pt. I	19.70 mean	18.40
AH5 Pt. II	19.66	17.11
AH5 Total	39.30	36.34
	d.d. = 9.62	s.d. = 9.11

$$t = 2.6 \text{ ** (for total AH5).}$$

The two groups were next examined with regard to interest in their studies. A major proportion of the 'work difficulty' group stated they had little or no interest in their work (Table 12).

Table 12

	Work difficulty	No work difficulty	
Interested	83	89	$\chi^2 = 25.74 \text{ ***}$ C.C. = 0.313
Uninterested	55	9	

The patients who complained of work difficulty were distributed diagnostically as shown in Table 13. These patients fell largely into the neurotic reaction, personality disorder and



affective psychoses groups which, on Ryan's (1960) test, did not differ significantly from each other. Each of these groups, however, split off rather markedly from the SCH and SX groups.

Table 13. *Comparisons between proportions with work difficulty in each diagnostic group*

Proportions of work difficulty complaints		Comparisons	Significance
ND	0.58	ND × SX	**
		ANX × SX	**
ANX	0.51	ND × SCH	**
		ND × P & Hy	n.s.
		ANX × SCH	*
		AFF × SX	*
AFF	0.49	ND × PD	n.s.
		ANX × P & Hy	n.s.
PD	0.48	AFF × SCH	*
		PD × SX	*
P & Hy	0.38	PD × SCH	n.s.
		P & Hy × SX	n.s.
SCH	0.19		
SX	0.18		

#### (5) *Suicides and attempted suicides*

Of the 25 patients who were referred after attempted suicide, two were diagnosed as being in a prodromal psychotic state, two as endogenous depression, three as cyclothymic, three as psychopathic, eight as reactive depression, four as personality disorders, two as having homosexual problems and one as an epileptic. The presenting symptom of fifteen of them was depression. Three other patients in the series were lost by suicide—all of whom were suffering from depressive illnesses; only one of these had made a previous attempt.

### DISCUSSION

The findings of the present study may perhaps best be seen against the general background of psychiatric services provided for undergraduates in Oxford. Not only has the medical staff of the Warneford Hospital long experience in dealing with the emotional and psychiatric problems of students but, in addition, the hospital provides an educational-psychological service which gives particular attention to study problems.

Comparison of the student admission rate with the Registrar-General's figures for the same sex, age group and social class shows the former rate to be about three times the national rate. At first sight this may suggest that the problem of student illness is even more serious than Parnell (1953) suggested. A higher referral rate must not, however, be taken as indicative of a higher incidence of psychiatric illness amongst Oxford students unless it is known that the criteria for referral are the same in each case. In a retrospective study it is not possible to obtain information regarding severity of symptoms on referral, but the fact that 4 per cent of students were found to have no psychiatric illness suggests that referral to the Warneford is not, *ipso facto*, evidence of a psychiatric disorder. Moreover, nearly a half of all referrals had 'work difficulty' as their presenting symptom. Whilst it may be argued that in a learning situation such as the academic one, psychiatric illness will first be manifested through under-functioning in study, it is also possible that some of these were primarily work problems. Indeed, students with the more serious illnesses, schizophrenia and affective disorders, complained of work difficulty significantly less frequently than those eventually diagnosed as neurotic disorders. Thus, though it is desirable that access to the psychiatric services of the Warneford Hospital be easy, for this very reason it is not possible to make direct comparisons of the hospital's referral rate with national rates.



The sample studied was a fairly heterogeneous one, not very different from any other psychiatric sample except for its restricted age range. The common nosological categories were represented and their distribution was similar to that in the general psychiatric population. The incidence of schizophrenia and affective illnesses in the students was not significantly different from that of non-students of the same age groups. The psychotics and personality disorders in general showed a greater frequency of morbid family history and early disorder, had a poorer prognosis, and did less well than other groups of patients. Thirty-four per cent of patients and only 12 per cent of controls had direct relatives who had been psychiatrically ill, thus indicating either an adverse genetic loading or the perpetuation of a neurotic family tradition in the former. Furthermore, 20 per cent of patients as against 2 per cent of controls had themselves been previously psychiatrically investigated. It thus appears that the psychological difficulties of many of these patients began before their entry to Oxford. In a sample of this age group it is not altogether surprising that developmental factors such as sexual difficulties, prompted referral. The fact that this group had a reasonably good prognosis suggests that in many the difficulties were transient.

On the basis of this study there is no evidence to suggest that there is any factor associated with the University as such which predisposes the students to psychiatric illness. The peak in the referral rate in the Hilary term (spring) is in accordance with what is known of the seasonal incidence of psychiatric illness in the general population (Mayer-Gross *et al.*, 1960; McCartney, 1961; Stengel *et al.*, 1958). Moreover it fails to offer supportive evidence for the concept of 'academic stress' (Spencer, 1958). On the basis of a stress hypothesis it would be expected that factors such as the impending crisis of final examinations would produce a peak incidence in the third term, or the difficulties of adjusting to a new social situation would result in a higher incidence in the first term. It might be argued that the peak in the Hilary term is itself a manifestation of 'academic stress', but if this were so it would be expected that the incidence in the 3rd year be greater than in the 1st and 2nd years. In fact, the peak incidence in the spring remains remarkably constant over the 3 years. Thus it seems more parsimonious to draw an analogy with the general psychiatric population.

Indications that social stresses are not necessarily associated with referral are given by the fact that grammar-school boys are not over-represented in the patient sample, although they were found to be socially more ill at ease than public-school boys (the proportions of the two groups being roughly the same in the population at risk).

Adverse home conditions were not observed more frequently among patients than controls, but a poor relationship with the parents was significantly associated with the patient group; the latter, however, is more likely to be a concomitant of the illness rather than a causal factor. But even were it to be regarded as an aetiological factor it is one not peculiar to students. For example, Andry (1960) indicates such a factor in delinquency.

The similarity between the MMPI profiles of patients and controls, although the former have higher scores on each scale, may be explained by the observation of Foulds (e.g. 1959) that this test measures both symptoms and personality traits. In terms of the distribution of personality variables the patients may not be very different from the population from which they are drawn, the two groups being discriminated from each other by symptoms. The fact that scores on the D-scale are the most significantly different seems to support this view (depression being the commonest presenting symptom).

Scholars are over-represented; this may be because tutors become concerned with their under-functioning more readily than with other students. No satisfactory explanation can be adduced to account for the under-representation of law students in the sample. The high referral rate for students reading Greats and English suggests that particular educational problems may be attached to these subjects. In fact, it is seen that these two groups complain of 'work difficulty' significantly more frequently than the rest of the sample (Greats  $\chi^2 = 4.10^*$  and English  $\chi^2 = 3.82^*$ ). Davidson and Hutt (in preparation) found that the Greats students were an intellectually able group but the English students scored significantly lower on AH5 than the rest of the patients ( $t = 3.47^{**}$ ). The over-representation of these two subjects suggests that they present primarily study problems. Two University lecturers, Finley (1963) and Hough (1963), have recently drawn attention to some of the difficulties which may be attached to reading these subjects at University. An analysis of what these problems may be would repay further investigation.



# REFERENCES

- ANDRY, R. G. (1960). *Delinquency and Parental Pathology*. London: Methuen.
- ANSTEY, E., DOWSE, R. F. & DUGUID, M. (1948). *Brit. J. educ. Psychol.* **18**, 156-160.
- BELL, C. R. (1962). *Brit. J. soc. clin. Psychol.* **1**, 81-93.
- DAVIDSON, M. A., LEE, D., PARNELL, R. W. & SPENCER, S. J. G. (1955). *J. ment. Sci.* **101**, 810-825.
- FINLEY, M. I. (1963). *The Sunday Times*, March 24, p. 27.
- FOULDS, G. A. & CAINE, T. M. (1959). *J. ment. Sci.* **105**, 183-189.
- HEIM, A. W. (1947). *Brit. J. Psychol.* **37**, 70-80.
- HOUGH, G. (1963). *The Sunday Times*, March 17, p. 26.
- HOWE, E. S. (1960). *J. appl. Psychol.* **44**, 115-120.
- MARTIN, R. M. & MARCUSE, F. L. (1958). *J. cons. Psychol.* **22**, 475-479.
- MAYER-GROSS, W., SLATER, ELIOT & ROTH, MARTIN (1960). *Clinical Psychiatry*. London: Cassell & Co., Ltd.
- MCCARTNEY, J. L. (1961). *Proceedings of the Third World Congress of Psychiatry*, Vol. II, pp. 1125-1129. University of Toronto/McGill University.
- PARNELL, R. W. (1951). *Lancet*, **1**, 731.
- PARNELL, R. W. (1953). *Times Educational Supplement*, April 3, p. 303.
- PAYNE, R. W. (1960). In *Handbook of Abnormal Psychology* (Ed. Eysenck), p. 193. London: Pitman Med. Pub. Co., Ltd.
- RYAN, T. A. (1960). *Psychol. Bull.* **57**, 318-328.
- SPENCER, S. J. G. (1958). *Lancet*, **ii**, 438-440.
- STENGEL, E., COOK, N. G. & KRUGER, I. S. (1958). *Attempted Suicide*. London: Maudsley Monograph, 4.
- WOLFE, D. (1954). *America's Resources of Specialised Talents: a Current Appraisal and a Look Ahead*. New York: Harper.

# ACKNOWLEDGEMENTS

During the period of this study one of us (C. H.) was in receipt of a research grant from the Oxford Regional Hospital Board. The patients were under the consultant care of Dr R. G. McInnes and Dr I. Skottowe. We are grateful to Dr S. J. G. Spencer for allowing us access to his records of the control subjects, to Dr C. Ounsted for many helpful suggestions, and to Mr H. C. A. Somerset for help and advice. We are indebted to Dr Alice Heim for permission to use her test prior to its publication.

*Manuscript received 17 August 1963*

## The Clinical Usefulness of some tests of Over-inclusive Thinking in Psychiatric Patients\*

By D. V. HAWKS

*Department of Psychology, Queen's University, Kingston, Ontario*

A retrospective study of fifty-eight psychiatric patients, who had been given three tests of over-inclusive thinking as part of their psychological examination, is reported. Clinical information relating to these patients was correlated with their test performance with a view to testing certain predictions made concerning the association of abnormal over-inclusion with specific symptoms and with clinical outcome. None of the expected relationships attained statistical significance.

In addition, the legitimacy of summing the three individual test scores to obtain a single composite over-inclusion score is considered and the conclusion arrived at that in their present form the tests used are not sufficiently reliable to recommend their clinical use.

### INTRODUCTION

Payne & Friedlander (1962), and Payne (1962), have suggested a number of objective measures of over-inclusive thinking in schizophrenia. Over-inclusion is defined, after Cameron (1938, 1939*a*, 1939*b*, 1944), as an inability to preserve conceptual boundaries as a result of which distantly associated, or even irrelevant, ideas come to be regarded as essential parts of the concept. In addition, speculations have been made with regard to the symptomatology which might be associated with this cognitive abnormality. Payne & Hewlett (1960) state, for instance, that it is 'tempting to relate overinclusion to paranoid delusions and ideas of reference, both of which can be regarded as an unwarranted generalization from the facts, patients with these symptoms being observed to perceive relationships where none exist and to take as evidence for their views data most people would not notice or judge to be irrelevant'. Incidental evidence for this view was furnished by their study: the two most over-inclusive individuals, as judged by test scores, were diagnosed as paranoid schizophrenics.

Some further indication of the plausibility of these hypotheses is provided by Payne, Caird & Lavery (1963). Fifteen schizophrenics with paranoid delusions, fifteen schizophrenics with no delusions and a control group of fifteen non-schizophrenic hospitalized patients were given the Benjamin Proverbs Test, the Mill Hill Vocabulary Scale and three of Babcock's Psychomotor Speed Tests. As they predicted, there was a significant positive relationship between the presence of delusions and over-inclusive thinking (assessed by the average number of words used in explaining the proverbs and two time scores).

\* This research was carried out while the author was the recipient of a Commonwealth Scholarship and was studying at the Institute of Psychiatry, London, England. Grateful acknowledgment is made of the assistance given by Dr M. B. Shapiro and Mr H. G. Jones, and to Dr R. W. Payne for furnishing the as yet unpublished manuscripts referred to and for his criticism of this paper.



The unpublished research of Craig (1960) has implications for these hypotheses. He wished to test the prediction that over-inclusiveness in schizophrenia is associated with the symptoms paranoia, delusions, ideas of reference and thought disorder. Over-inclusion was defined as an expansion of the conceptual boundaries, and measures based on tests of concept formation were expected to intercorrelate in a way which would permit the identification of such a dimension. A second dimension of retardation was expected to underlie the intercorrelations between tests of mental speed. Sixty-six patients were examined for eleven symptoms which were rated by the doctor in charge of the case. In addition, the patient underwent a psychological examination, which involved the use of thirteen tests from which were derived sixty-two measures. Thirty-eight of these measures were subsequently intercorrelated and subjected to a principal components analysis. Five principal components were extracted and three rotated 'to produce meaningful structure'—the rotated components being then referred to as 'factors'. The first factor Craig called 'over-inclusion', the second 'poor concept formation' and the third 'conceptual retardation'.

Factor scores were then calculated for each of the sixty-six patients and correlated with the presence or absence of each symptom so as to produce regression equations for predicting each symptom jointly from the factors. The method of analysis provided the regression coefficients, or weights, applicable to each factor or component for the best possible discrimination between those who had a particular symptom and those who did not. The significance of the discrimination was tested by an analysis of variance.

The first factor was found to provide the best discrimination between those patients who had, and those patients who did not have, delusions; and between those patients who had/did not have thought disorder; the second factor the best discrimination of paranoia, thought disorder and, to a lesser extent, delusions; while scores on the third factor identified obsessive-compulsive symptoms; and low scores on the fourth component identified depressives. Ideas of reference were not distinguished by any test or factor.

The factor scores were not found to differentiate between the diagnostic groups, schizophrenia and depression, but over-inclusion, defined as the presence of the symptoms of delusion or thought disorder, and retardation, defined as the presence of the symptoms of depression or slowness, could be significantly differentiated by the second factor.

The question thus arises as to whether it might be possible to obtain an objective measure of over-inclusion which has implications for symptomatology and prognosis in the manner claimed by Payne & Hewlett (1960) and Payne & Friedlander (1962). A weakness of all the published studies quoted is that they do not consider the reliability of the tests used, nor the legitimacy of combining such tests into a single battery, though the consistency of the findings from some of the tests would suggest they have some stability. Such indirect evidence of their reliability cannot, however, be taken as sufficient.

In addition, then, to considering the possible symptomatic correlates of over-inclusion the present investigation sought to determine the reliability of the tests used in its measurement.

## METHOD

### *The population*

A retrospective study of fifty-eight patients given three tests of over-inclusion while patients of the Maudsley and Bethlam Royal Hospitals, London, between the years 1959 and 1962 was undertaken. None of Payne & Hewlett's (1960), nor Craig's (1960), subjects were included in this sample as the present study followed these investigations.

Such patients were almost invariably referred for psychological testing for the reason that thought disorder was suspected (a majority of such cases would undoubtedly be patients whose diagnosis was in doubt). As such, they do not represent a 'pure' schizophrenic group as used by Payne & Hewlett (1960) in their investigation. This difference excepted, the present population closely resembled in age, I.Q. and mean over-inclusion and retardation scores the schizophrenic sample of Payne and Hewlett. With respect to age and I.Q., the present population also closely approximates that employed by Payne, Caird & Laverty (1963) and Craig (1960).

The similarity of population characteristics was partly achieved by excluding six patients who fell outside the age and I.Q. ranges employed by Payne & Hewlett (1960). In addition to these six, two other patients were excluded because the psychologist who did the testing



Table 1. *Statistical characteristics of total patient population with respect to intelligence, age and over-inclusion\**

(The comparable values for Payne & Hewlett's (1960) schizophrenic group (N = 20) are given in brackets).

	I.Q.	Age (years)	Over-inclusion
M	104.9 (101.26)	30.79 (32.10)	10.19 (12.20)
S.D.	16.4 (12.54)	9.50 (8.32)	5.73 (6.18)
V	270.4 (157.22)	90.61 (69.26)	32.84 (38.27)
Range	61-132 (80-126)	17-57 (18-48)	3-34 (2-25)
N	52	58	58

\* The number of entries involved in each of the computations varies because not all the data was available for every patient.

reported a lack of confidence in the results obtained, and eight were excluded because information relating to their clinical state was inadequate.

Examination of the case summaries of the patients selected for inclusion in the present study showed that thirty-two received the final diagnosis of schizophrenia ('schizophrenic group')—of these, eleven were labelled paranoid schizophrenics; while twenty-six received other diagnoses ('other diagnoses group'). Of the latter group, the greatest proportion were labelled 'schizo-affective' (35 per cent), next came 'personality disorder' (20 per cent)—the remainder represented a variety of conditions. There were thirty-five males and twenty-three females.

#### *The test battery*

The tests used in the present investigation were the Goldstein-Scheerer Object Sorting Test (Goldstein & Scheerer, 1941), the Benjamin Proverbs Test (Benjamin, 1946) and the Shaw Block Test (Payne & Hewlett 1960; Bromley 1956).

Previous research (Payne & Hewlett, 1960; Payne, Mattussek & George, 1959; Craig, 1960; Payne, Caird & Lavery, 1963) has shown these tests differentiate significantly between patients diagnosed as schizophrenic and patients diagnosed alternatively as neurotic or depressive.

In addition, there is evidence that these tests are relatively independent of differences in intelligence (Payne & Hewlett, 1960), at least within the range of intelligence of patients included in the sample; further, slowness or retardation has been shown to be an important variable in their performance (Payne & Hewlett, 1960).

The tests have been described in detail by Payne and Hewlett; their administration and scoring followed that set down by these authors, a patient's combined over-inclusion score being obtained by summing the transformed scores of the individual raw scores.

The present battery is similar to that described by Payne & Friedlander (1962). However, these authors used the Payne Object Classification Test instead of the Shaw Block Test and the unusualness score of the Goldstein-Scheerer Object Sorting Test. In terms of factor saturation and discrimination between the diagnostic groups, however, there appears to be little basis for choosing between these two alternative combinations of tests.

Where possible, the Wechsler-Bellevue (Wechsler, 1944) or W.A.I.S. (Wechsler, 1955) Verbal Scale I.Q. were used as measures of I.Q. In those cases where no Wechsler I.Q. was available, a comparable I.Q. was derived from the Mill Hill Vocabulary Scale (Raven, 1948).

#### *The clinical variables*

The clinical variables selected included those which have been shown to have significant associations with over-inclusion as objectively defined and which were estimated to occur frequently enough in the population under consideration to justify their statistical analysis,



together with a number of others thought by the author to be relevant. They comprised the following:

- Diagnosis;
- Sub-diagnosis;
- Delusions;
- Hallucinations;
- Ideas of reference;
- Obsessive-compulsions;
- Orientation;
- Condition on discharge.

Information relating to these variables was obtained from each patient's Case Summary. The 'diagnosis' and 'sub-diagnosis' recorded for each patient was arrived at during his presentation to a case conference involving all the psychiatrists working on the patient's ward.

Information was sought about the presence or absence *at admission* to the hospital of the symptoms—delusion, hallucination, ideas of reference and obsessive-compulsions. It was thought that observations made at this time related to the condition of the patient at the time psychological testing was most probably requested. This time factor was considered important, since it has been suggested by Payne, Anchevich & Lavery (1963) that over-inclusion is a symptom of an illness and not a permanent feature of a person who at some time is diagnosed schizophrenic. There is, even so, no certainty that the assessment of the clinical variables coincided with the assessment of over-inclusion, an uncertainty which results from the retrospective nature of the investigation.

With regard to the variables 'orientation' and 'condition on discharge', the following conventions were adopted. Orientation was recorded as correct, partial or lacking. Condition on discharge was recorded as largely unchanged ('unchanged', 'slightly improved') or improved ('improved', 'moderate improvement', 'marked improvement', 'symptom free'). General uniformity, in the doctors' recording of the above symptoms, was observed in the majority of cases—only obsessive-compulsive symptoms and ideas of reference were not reported in a considerable number of cases.

Where no information was given about the presence or absence of one of the symptoms this was noted. The experimenter was not aware of the patient's over-inclusion score while recording the clinical information relating to each patient. The doctor in charge of the case was aware of the psychological test results when he wrote the case summary. To this extent, then, the comparisons are contaminated, though subsequent analysis did not show this to have affected the results.

## RESULTS

### *Over-inclusion*

Examination of the over-inclusion histogram for the total group (see Figure 1) shows that of the sixteen patients getting scores of 12 and above all but six were diagnosed schizophrenic—a finding which tends to confirm the conclusion that over-inclusion is a characteristic of schizophrenia. The comparable data from Payne and Hewlett's study is that, of fifteen patients getting scores 12 or above, all but five were diagnosed schizophrenic. It is obvious from Figure 1, however, that within the schizophrenic sample there is a wide dispersion of scores.

Of special note is the fact that in both studies approximately 30 per cent of the schizophrenics tested received scores greater than 13, the score regarded by Payne and Hewlett as separating off those patients showing 'abnormal over-inclusion'. This coincidence tends to confirm the hypothesis that of a group of acute schizophrenics from one-third to one-half will be shown to be abnormally over-inclusive when compared with other psychiatric populations, or with normals (Payne, 1960).

### *Over-inclusion and diagnosis*

Aside from the diagnosis 'schizophrenic', no diagnostic category was sufficiently represented in the test population to enable a statistical comparison of various diagnostic groups to be made. In considering the association of over-inclusion with diagnosis, then, only two categories were employed—'schizophrenic' versus 'other diagnoses'.

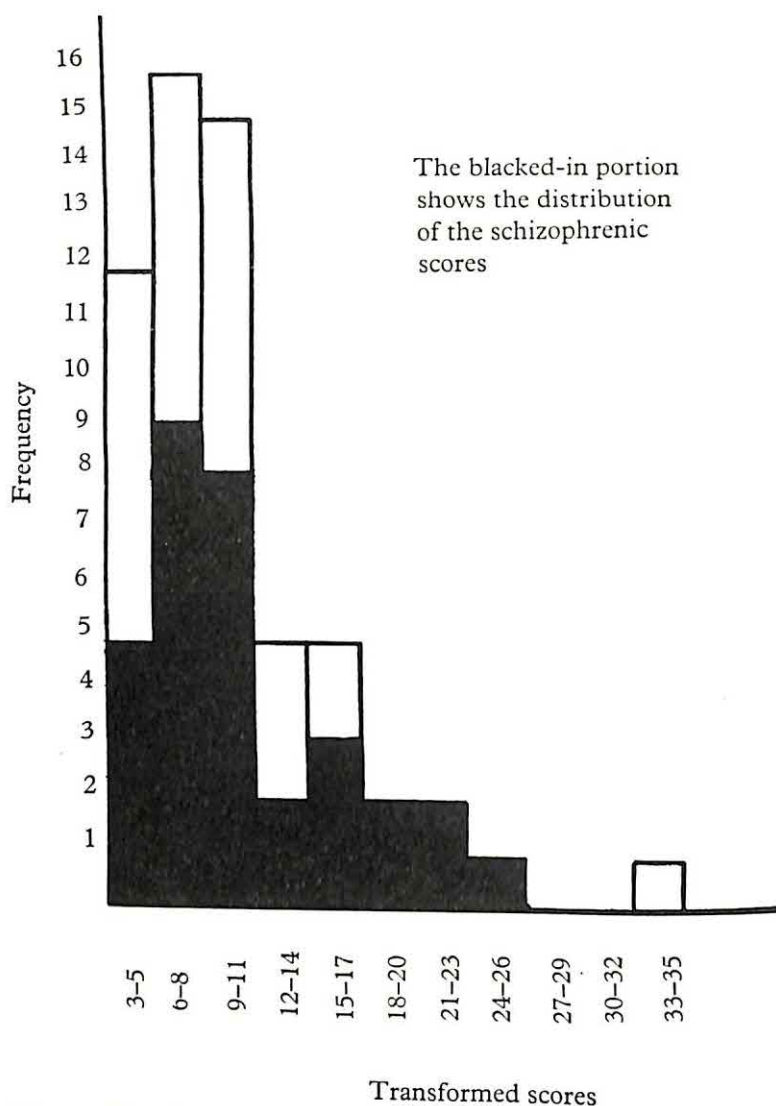


Fig. 1. Distribution of composite over-inclusion scores for total patient population (N = 59).

The fact that the present schizophrenic sample was slightly less over-inclusive on the average than Payne & Hewlett's (1960) group was thought to be possibly due to it containing some cases of doubtful diagnosis. Nevertheless, it was expected that within the present population patients diagnosed as schizophrenic would be more over-inclusive than patients not so diagnosed. In fact, it was found that the schizophrenics in the total sample did not, on the average, score higher on the tests of over-inclusion than patients having other diagnoses (see Table 2).

Table 2. Comparison of schizophrenics versus 'other diagnoses' on combined over-inclusion score

	Schizophrenics	'Other diagnoses'
M	10.80	8.96
S.D.	5.58	4.02
Range	3-26	3-20
N	32	26

$$t = 1.446 \text{ N.S.}$$



Within the group studied (patients referred because of suspected thought disorder) there does not appear to be a greater tendency for schizophrenics to get higher scores on the over-inclusion battery than patients having other diagnoses. This is true in spite of the fact that, of the sixteen patients getting scores on the over-inclusion battery of 12 or higher, all but six were diagnosed schizophrenic. This finding highlights the fact that, while schizophrenics as a rule get the highest scores on tests of over-inclusion, this characteristic is by no means typical of all schizophrenics, some getting very low scores. This would suggest that statements to the effect that 'abnormal over-inclusion is a particular characteristic of schizophrenics' (Cameron, 1944; Payne, 1962) need qualification.

The suggestion that it is paranoid schizophrenics in particular who get high scores on over-inclusion tests was tested by comparing the average score obtained by the paranoid schizophrenics included in the sample ( $N = 11$ ) with that of the remainder of the schizophrenic group ( $N = 21$ ). These two groups were not matched for age, intelligence or sex (although the evidence of the present study suggests that intelligence and sex do not correlate significantly with scores on the over-inclusion tests).

The difference in mean score for these two groups was not found to be statistically significant (see Table 3).

Table 3. *Comparison of undifferentiated schizophrenics versus paranoid schizophrenics on combined over-inclusion score*

	Undifferentiated schizophrenics	Paranoid schizophrenics
M	11.14	10.27
S.D.	5.84	5.27
Range	3-26	5-21
N	21	11

$$t = -0.414 \text{ N.S.}$$

#### *Over-inclusion and symptomatology*

With the exception of 'orientation', 'obsessive compulsions' and 'hallucinations', the clinical variables included in the present study have been shown by previous research to be correlated with abnormal over-inclusion. Orientation was included as one of the clinical variables in the present investigation since it was felt that high scores on the over-inclusion tests may result from general disorientation (failure to remember instructions, etc.) which may be present even in the absence of clinically recognizable thought disorder. In fact, only a very small minority of the patients included in the analysis showed any degree of disorientation. While insufficient numbers of disorientated patients were thus available for a systematic test of the above hypothesis to be made, the implication is that high scores on the over-inclusion battery are unlikely to be merely a reflection of general disorientation. A collateral conclusion is that some patients clinically judged to be orientated can manifest thought disorder of the type measured by the over-inclusion battery.

Obsessive-compulsive symptoms were included among the clinical variables because Craig found these to be associated with low scores on tests of over-inclusion, and also because of the present author's notion that, contrary to Craig's finding, the presence of obsessive-compulsive symptoms would lead to high scores being obtained on the tests of over-inclusion, by virtue of the obsessive's 'attention to irrelevant details, over-meticulous circumstantiality, vacillation and excessive partially relevant and irrelevant verbiage—the coming into consciousness of many associations related to an impulse or task, with concomitant difficulty and doubt in selecting the task-appropriate one' (Rapaport, 1951). It was hoped that the present investigation would shed some light on these contrary expectations. Hallucinations were included among the clinical variables out of interest, rather than because there was any definitive evidence from previous research to suggest an association with abnormal over-inclusion.

The predicted relationships between over-inclusion scores and the symptoms of delusion, hallucination, ideas of reference and obsessive-compulsions, were tested employing the total group, by calculating chi-squares using as the cut-off point for over-inclusion the mean of



the combined over-inclusion scores for the total group and presence or absence as the two categories of each of the symptoms.

None of the predicted associations attained statistical significance (see Table 4).

Table 4. *Chi-square tests of association between combined over-inclusion score and clinical variables*

(Total patient population)

Comparison	N	$\chi^2$	P
Over-inclusion $\times$ Delusion	58	0.11	N.S.
Over-inclusion $\times$ Hallucination	56	0.0056	N.S.
Over-inclusion $\times$ Obsessive compulsions	42	0.0029	N.S.
Over-inclusion $\times$ Ideas of reference	22	0.14	N.S.

In order to obtain data more relevant to the hypotheses put forward by Payne & Hewlett (1960), the same was done for the schizophrenic group only ( $N = 32$ ). None of the expected associations were borne out (see Table 5).

Table 5. *Chi-square tests of association between combined over-inclusion score and clinical variables*

(Schizophrenic group only)

Comparison	N	$\chi^2$	P
Over-inclusion $\times$ Delusions	32	0.034	N.S.
Over-inclusion $\times$ Hallucinations	30	0.051	N.S.
Over-inclusion $\times$ Obsessive compulsions	23	0.017	N.S.
Over-inclusion $\times$ Ideas of reference	14	0.110	N.S.

In order to rule out the possibility that freak scores on the over-inclusion battery exerted a disproportionate effect on the analysis, a second set of calculations was carried out using as the cut-off point for over-inclusion the median value for the schizophrenic group. None of the predicted associations attained statistical significance. The mean would appear the more justifiable value to use as it more nearly corresponds with the score regarded by Payne & Hewlett (1960) as dividing those patients showing abnormal over-inclusion (when compared with controls) from those lacking this characteristic. The median value for the schizophrenic group falls lower in the range of scores.

In addition to calculating chi-squares for the above variables, product moment correlations were also computed for each of the clinical variables and over-inclusion. This was achieved by regarding the presence of the clinical variable under consideration as a score of one and its absence as a score of zero. This was done to show that the insignificant relationships revealed by the chi-square analyses were not the result of information being lost from the analyses by virtue of considering scores on the over-inclusion battery as being only above or below the mean. None of the correlations was shown to be significant.

A criticism which can be made is that the statistical analyses carried out have been concerned only with the symptoms considered individually, the supposition being that if the symptoms were considered in combination a significant association would be found.

While the present design does not permit a direct test of this hypothesis, an indication of its plausibility can be obtained by forming groups from the patients with two or more of the clinical symptoms supposed to be associated with over-inclusion (delusions, hallucinations, ideas of reference) and from the patients showing less than this number of symptoms. When this is done, using a chi-square test and taking as the cut-off point for over-inclusion the mean for the total group, no significant relationship between the number of clinical features and over-inclusion is found ( $\chi^2 = 0.0075$ ). This would suggest, at least, that the failure of the present design to consider symptoms in combination is not a serious deficiency or likely to account for the non-significant results obtained.



*Over-inclusion and prognosis*

Information relating to 'condition on discharge' was sought because of the suggestion made by Payne, Friedlander, Lavery & Haden (1963) that the prognosis for schizophrenics demonstrating over-inclusion may be good. While no long-term follow-up was attempted in the present investigation, it was felt that if the hypothesis was correct, then the good prognosis of over-inclusive patients would be reflected in their condition on discharge.

No consideration was given to the length of hospitalization or to the mode of treatment, the assumption being made that while these two variables were most probably related to prognosis they were not likely to be affected by the score a patient received on the over-inclusion tests. (The assumption of independence of the two variables correlated is therefore not violated.)

The predicted relationship between the combined score on the over-inclusion tests and prognosis was tested, employing the total group, by calculating chi-square using as the cut-off point for over-inclusion the mean of the total group and 'largely unchanged' versus 'improved' as the categories of condition on discharge. The chi-square value so obtained was not significant ( $\chi^2 = 0.39$  with an N of 50). Employing only the schizophrenic group, the value was 0.34 with an N of 27. Using the median value as the cut-off point for over-inclusion did not affect the results.

*Test reliability*

The four measures included in the present battery were selected because they have been shown to differentiate (in the direction required by the theory) patients diagnosed as schizophrenic and patients given other diagnoses. No consideration has been given in previous studies to the legitimacy of combining the tests' results into a single score.

In order to test this procedure, the intercorrelations between the individual tests, and their correlation with the total over-inclusion score, were calculated. Insignificant correlations, or widely varying correlations, would suggest that the subtests are measuring different things, in which case the correlation of symptoms with over-inclusion would be more justifiably carried out with regard to each of the subtest scores, rather than with the composite score, as has been done in the present investigation.

Calculation of the correlation of each of the subtests with the total, using the data of the present investigation, produced the values shown in Table 6—the comparable values computed from Payne & Hewlett's (1960) data are shown in brackets. The fact that only one of the four correlations computed from the present data attains significance at the 5 per cent confidence level suggests that the combination of the four measures into a single composite score is unjustifiable when the present population is considered. The correlations obtained from Payne and Hewlett's data, while a little higher, would still seem too low to recommend the battery's clinical use.

Table 6. *Correlation of the four subtest scores with the total score—corrected for contamination*

(Comparable figures from Payne & Hewlett (1960) are shown in brackets)

Subtest	Correlation with total score	
S	0.170	(0.329)*
O	0.249	(0.468)*
U	0.310*	(0.510)*
P	0.239	(0.449)*

\* Significant at 0.05 level of significance.

S. Shaws blocks C + D responses.

O. Goldstein-Scheerer O.S.T. average number of objects handed over.

U. Goldstein-Scheerer O.S.T. number unusual sortings.

P. Benjamin Proverbs test average number of words used.

The intercorrelations of the four subtests are shown in Table 7. The comparable values computed from Payne & Hewlett's (1960) data are shown in Table 8.

Table 7. *Intercorrelations of the four subtests making up the over-inclusion battery used in the investigation*

	(N = 58)			
	S	O	U	P
S	—	0.146	0.047	0.196
O		—	0.303*	-0.030
U			—	0.307*
P				—

\* Significant at the 0.05 level of confidence.

Table 8. *Intercorrelations of the measures combined in the battery of over-inclusion tests*

	(Data from Payne & Hewlett, 1960)				
	C	D	O	U	P
C	—	0.326*	0.072	0.276*	0.229
D		—	0.371*	0.279*	0.256*
O			—	0.456*	0.494*
U				—	0.500*
P					—

\* Significant at the 0.05 level of confidence.

C. Number of C responses given in the Shaw Block test.

D. Number of D responses given in the Shaw Block test.

O., U., P. As defined on page 193.

Estimating the reliability of the test battery from the average correlation of the four subtest scores with the total score (Guildford, 1956), a procedure which assumes that the subtests are relatively homogeneous in respect of what they are measuring (an assumption not fully justified in the present case), reveals that this value is not much higher than 0.2 when the present data are used, and is about 0.5 when Payne & Hewlett's (1960) data are used. Even supposing that the estimate obtained from Payne and Hewlett's data is the more accurate, this figure would seem to preclude the battery's clinical use. The experimenters who have employed the above tests as measures of overinclusion do not report their reliability. The present evidence suggests that these reliabilities, if calculated, would be low.

## CONCLUSIONS

The conclusions which seem to follow from the present investigation are, first, that there is no evidence that the combined score on the tests of over-inclusion has implications for the following variables—diagnosis, the presence or absence of the symptoms of delusion, hallucination, ideas of reference, obsessive-compulsions (as these are defined by the operations of the study). While it is true that a high score probably indicates a diagnosis of schizophrenia, a low score does not rule out the possibility of schizophrenia. The combined score on these over-inclusion tests has no implication for immediate prognosis as reflected in condition on discharge.

Second, only negligible intercorrelations were found between the three measures of over-inclusion. There is, therefore, no real basis for the practice of combining the tests into a battery and applying such clinically without first investigating the reliability and validity of the battery for the population in question.

The present sample differs from the clear cut cases included in Payne & Hewlett's (1960) investigation because it was made up of patients referred for psychological



testing, presumably because of some doubt as to their diagnosis. It is conceivable that the failure in the present investigation to demonstrate the expected relations is attributable to this difference in populations. In defence of the present approach, it may be argued that the validation of a test battery for clinical use is more appropriate in the population to whom it will be administered in practice than in a population whose members are not, in fact, referred to the psychologist for testing.

## REFERENCES

- BENJAMIN, J. D. (1944). A method for distinguishing and evaluating formal thinking disorders in schizophrenia. In *Language and Thought in Schizophrenia* (ed. J. S. Kasanin). Berk. & Los Angeles: University of California Press.
- BROMLEY, D. B. (1956). Some experimental tests of the effect of age on creative intellectual output. *J. Geront.* **11**, 74-82.
- CAMERON, N. (1938). *Reasoning, Regression and Communication in Schizophrenics*. Washington: Psychol. Mong. Soc.
- CAMERON, N. (1939a). Schizophrenic thinking in a problem solving situation. *J. ment. Sci.* **85**, 1012-1035.
- CAMERON, N. (1939b). Deterioration and regression in schizophrenic thinking. *J. abnorm. (soc.) Psychol.* **34**, 265-270.
- CAMERON, N. (1944). Experimental analysis of schizophrenic thinking. In *Language and Thought in Schizophrenia* (ed. J. S. Kasanin), pp. 50-64.
- CRAIG, W. J. (1960). An experimental investigation of certain thought disorders in abnormal mental states. Unpub. Ph.D. Thesis, University of London.
- GOLDSTEIN, K. & SCHEERER, M. (1941). Abstract and concrete behaviour: an experimental study with special tests. *Psychol. Monogr.* **53**, No. 2.
- GUILDFORD, J. P. (1956). *Fundamental Statistics in Psychology and Education* (3rd ed.). New York: McGraw Hill Book Co.
- PAYNE, R. W. (1960). Cognitive abnormalities. In *Handbook of Abnormal Psychology* (Ed. H. J. Eysenck). London: Pitman.
- PAYNE, R. W. (1962). An object classification test as a measure of over-inclusive thinking in schizophrenic patients. *Brit. J. soc. clin. Psychol.* **1**, 213-221.
- PAYNE, R. W., MATTUSSEK, P. & GEORGE, E. I. (1959). An experimental study of schizophrenic thought disorder. *J. ment. Sci.* **105**, 627-652.
- PAYNE, R. W. & HEWLETT, J. H. G. (1960). Thought disorder in psychotic patients. In *Experiments in Personality*, Vol. II, (Ed. H. J. Eysenck). London: Routledge & Kegan Paul.
- PAYNE, R. W. & FRIEDLANDER, D. (1962). A short battery of simple tests for measuring over-inclusive thinking. *J. ment. Sci.* **108**, 362-367.
- PAYNE, R. W., ANCEVICH, S. S. & LAVERTY, S. G. (1963). Over-inclusive thinking in symptom-free schizophrenics. *Canad. J. Psychiat.* (in press).
- PAYNE, R. W., CAIRD, W. F. & LAVERTY, S. G. (1963). Over-inclusive thinking and delusions in schizophrenic patients. *J. abnorm. (soc.) Psychol.* (in press).
- PAYNE, R. W., FRIEDLANDER, D., LAVETRY, S. G. & HADEN, P. (1963). Over-inclusive thought disorder in hospitalized chronic schizophrenic patients and its response to proketazone. *Brit. J. Psychiat.* **109**, 523-530.
- RAPAPORT, D. (1951). *Organization and Pathology of Thought*, Chap. 26. New York: Columbia University Press.
- RAVEN, J. C. (1948). *A Guide to Using the Mill Hill Vocabulary Scale with Progressive Matrices* (1938). London: H. K. Lewis.
- WECHSLER, D. (1944). *The Measurement of Adult Intelligence*. Baltimore: Williams & Williams Co.
- WECHSLER, D. (1955). *Manual Wechsler Adult Intelligence Scale*. New York: Psych. Corp.

Manuscript received 26 April 1963

Revised manuscript received 10 December 1963



## Some Moral Concepts and Judgments of Junior School Children\*

By M. L. KELLMER PRINGLE† *National Bureau for Co-operation in Child Care*

AND

J. B. EDWARDS† *North Wales Child Guidance Clinics*

The following three devices were used with a sample of 226 children: (1) The 'Ideal Person' Test, where the child names the person he would most wish to be like and gives reasons for his choice; (2) The 'Moral Wickedness' Test, where the subject describes what he considers to be the most 'wicked' deeds; (3) The 'Moral Incidents' Test, where ten hypothetical situations, involving honesty, loyalty, responsibility and punishment, are presented together with alternative methods of meeting them. The results were analysed in relation to sex and intelligence as assessed by the Stanford-Binet Scale.

The majority of children chose as their ideal characters personally unknown to them. Both sexes tended to choose male figures but boys mentioned a relatively greater proportion of characters associated with war, adventure, politics and sport, whereas girls chose a higher proportion of people distinguished because of moral or religious qualities. Less able children more often named as their ideal someone from their immediate circle of relatives or friends and gave fewer reasons for their choice than the more intelligent group.

The most frequently mentioned 'wicked deeds' were murder, physical cruelty, stealing and cruelty to animals. Though the sexes showed considerable unanimity in their list of wicked actions, a higher proportion of girls than boys considered stealing, cruelty to animals, lying and swearing to be wicked. The most intelligent group produced a greater number and variety of morally wicked actions.

The findings from the 'Moral Incidents' Test suggest that straightforward moral issues are readily understood by children but that the duller child's moral judgment is prone to become confused or blurred in the face of complexity or subtlety.

### I. INTRODUCTION

Compared with the interest shown in children's intellectual and social growth, there is a paucity of research in this country on their moral development, particularly during the junior school period. In one of the earliest systematic studies McGrath (1923) used a variety of quite ingenious techniques, covered a wide age range (from 6 to 18 years) and her sample was large, even by present-day standards (over 4000 children attending six different schools). Its main aim was to seek information on which could be based 'a course in moral instruction, so widely heralded and so badly needed'. Though many of the conclusions strike one now as over-simplified, the monograph contains much that is suggestive, such as the attempt to outline the stages of moral development and the growth of moral principles between the ages of 6 to 18 years. Macaulay & Watkins (1926) studied the effects of environmental and other influences on moral values, together with the changes that take place with increasing age.

Perhaps the best known and most extensive study of value-judgments was the 'Character Education Inquiry' by Hartshorne & May (1930). From the results of

\* The analysis of the data was made possible by grants from the Mental Health Research Fund and the Noel Buxton Trust.

† Formerly of the Department of Child Study, University of Birmingham.



over 100 tests, scales and real life situations, the authors concluded that moral knowledge and moral behaviour were not closely related among their adolescent population. Their findings were much discussed and criticized, particularly their view of the specificity of moral conduct (Maller, 1934; Allport, 1937; Eysenck, 1953).

Piaget's (1932) classic research into children's moral judgment led to a series of studies (Harrower, 1934; Lerner, 1937; MacRae, 1950, 1954; Havighurst & Taba, 1949; Havighurst & Neugarten, 1955; Morris, 1958); however, only the first and last deal with English children. By studying the ways in which children of various ages play marbles, Piaget developed his theory of how children learn social or moral norms. He distinguished two main stages: during the first, rules are regarded as being absolute, unalterable and externally determined; this Piaget calls the transcendental stage or absolute orientation to norms. Punishment is also viewed in an 'absolutistic' way since young children tend to think that it should be meted out according to the enormity of the crime, regardless of intent or responsibility, i.e. the child who accidentally breaks ten plates should be punished more severely than one who breaks only one. Later, by about the age of 11 or 12 years, constraint is replaced by reciprocity and co-operation, rules are seen as useful conventions which are maintained, but can also be changed, by mutual agreement; this maturer and more realistic attitude to rules, Piaget called the autonomous stage or relative orientation to norms. By this age, too, children have learned that punishment should be related to intent, provocation or mitigating circumstances.

Harrower (1934) repeated Piaget's experiments on the child's conception of 'punishment' and 'cheating' with two aims in mind: first, to investigate whether or not different results would be obtained with children in this country; and secondly, 'by carrying out the experiment with two different social groups, to discover whether or not Piaget has minimized the environmental factor'. The group of children who were socially equivalent to Piaget's (namely those from the poorer parts of London) responded in a very similar manner; however, the responses of her control group, consisting of children from cultured, well-to-do homes, differed both quantitatively and qualitatively. Harrower concluded that 'either the stages of development which Piaget has been emphasizing are not a universal characteristic of development *per se*' or that in certain environments these stages can be much accelerated.

Lerner (1937), studying the relationship between social status, parental authority and moral judgment among 6- to 12-year-olds, also found that children of high-status parents are less liable to see moral principles as externally imposed and unvaryingly rigid. In later work with adolescents (MacRae, 1950, 1954) criticism was levelled against Piaget's concept of 'autonomy' on the grounds that it included many different and distinct aspects of moral development, not all of which were highly correlated. Morris (1958) studied the relationship between age and moral development during adolescence. He found marked 'situational differences' in value judgments, changes in the direction of greater autonomy with increasing age and marked discrepancies between what adolescents thought 'should be done' and 'would actually be done' in various problem situations; sex differences were slight but greater conformity was found in the modern school.



Thus Piaget is being criticized for his almost exclusive stress on maturation in bringing about the transition from one stage of morality to the next with a consequent neglect of social, cultural and intellectual influences as well as the impact of differential child-rearing practices and individual differences among children (Havighurst & Taba, 1949; Havighurst & Neugarten, 1955; Eysenck, 1960). Peters (1960) sums up the present position by saying that Piaget has provided a useful framework for research but many of his concepts will need clarifying; the studies of others are suggestive but so far inconclusive. His own conclusion is that 'in this ill-explored field of moral development what is needed more than in almost any other field in psychology is a combination of concrete investigations with conceptual clarity'.

A recent study by Bray (1962) merits discussion in detail since it deals with a group of children similar to ours in age and general composition (i.e. English and prior to the 11-plus selection). As part of the selection examination they were asked to write a composition on an admired person. As this was set by their own teacher and was part of an examination, widely regarded as all-important, it is likely that the choice of person was influenced by what the children thought their teachers would approve of, rather than by their own interests and feelings. The main findings were that 'remote' choices predominated over 'local' ones, that boys mainly chose characters of their own sex, whereas girls tended to do so for 'local' persons while for 'remote' characters only half their choices were women; that recognition and success were the predominant features of admired characters while the giving and receiving of help and affection was more frequently mentioned by girls than boys; the latter often described characters associated with war, politics, explorations and sport, whereas girls chose a greater number of humanitarian and religious figures. Bray considers that adequate provision is needed 'of potential heroes and heroines of like sex to the pupils and according to different appeals as between the two sexes'. However, he feels education should not accentuate sex differentiation unduly but instead 'a wide range of both male and female characters in diverse fields of human endeavour should be presented to both boys and girls with the aim of promoting broad sympathies'.

Since the measurement of moral behaviour or conduct is beset by so many difficulties, Edwards (1959) investigated certain moral attitudes held by secondary modern school boys in their first and last year at school. Again, many individual differences were found and older boys showed as much uncertainty as younger ones when considering moral conflicts or defining moral concepts; very rarely were outstanding moral or religious figures named as 'ideal persons' and the influence of film and television personalities seemed to be rather neutral.

In the study to be described here, some of the instruments devised and used in the above mentioned investigation were employed (with modifications where necessary to suit younger children).

## II. SUBJECTS AND METHODS

This investigation forms part of a larger, longitudinal project concerned with various aspects of children's intellectual, emotional and educational development during the junior school



years. The 226 children who took part in this study comprised the entire fourth year in two junior schools in the Midlands; there were 109 boys and 117 girls. Both schools are situated in areas populated mostly by skilled working-class families, many of whom live in owner-occupied houses; the proportion of professional, managerial and business people is small as is the number of semi-skilled and unskilled workers; a fair proportion of families live in Council houses, some of which are on post-war estates.

The following three devices were used: (1) the 'Ideal Person' Test, in which the child was asked to name the person he would most wish to be like and to give reasons for his choice; (2) the 'Moral Wickedness' Test, where the child was invited to describe what he considers to be the most 'wicked' deeds or actions, listing them in descending order of wickedness; to be the most 'wicked' deeds or actions, listing them in descending order of wickedness; the questions used here were those formulated by Macaulay & Watkins (1926); (3) the 'Moral Incidents' Test, where ten hypothetical situations, involving honesty, loyalty, responsibility and punishment, were presented and the child was asked to choose between three alternative methods of meeting or reacting to each situation. Details of this test will be given in the next section, together with the children's replies.

The mean age of the total sample was 11 years (ranging from 10 years 6 months to 11 years 6 months) and the mean intelligence quotient on the Terman-Merrill Intelligence Scale (1937) Form L was 112.1 (ranging from 74-168,  $\sigma = 17.31$ ); the Roberts-Mellone corrections were used where applicable. To explore whether there were differences in moral concepts and judgments at various levels of intelligence, the sample was divided into three groups:

Able children (I.Q. 121 to 168)	N = 77
Average ability (I.Q. 95 to 120)	N = 109
Low average to dull (I.Q. 74 to 94)	N = 40

The mean I.Q.s of the three groups were 131.2, 108.1 and 86.5 respectively. The class teachers were not present while the tests were administered. In many other respects the procedure followed was similar to that employed over the 4 years of the larger, longitudinal study: the children were assured that their answers would not be seen by their headmasters, class teachers, 'the office', etc., and that their work had no relationship or influence upon the 11-plus examination; that the investigation aimed at finding out children's opinions on these questions and that there were no right or wrong answers. In addition, they were told that these particular questions were not timed, so that they could work at their own speed without hurrying; that spelling did not matter and that they could ask how to spell any word they did not know. With the most backward forms, the 'Moral Incidents' Test was done with the children to overcome the difficulties of poor readers; those who could manage, were allowed to carry on by themselves. It was stressed that anyone, man or woman, could be named as the 'ideal person', and that the list of wicked deeds could be as long as each child wished to make it; also that they could give as many reasons as they could think of, for considering these actions to be wicked.

### III. RESULTS

#### 1. The 'Ideal Person' Test

As can be seen from Table I,\* personally unknown or 'remote' figures predominated for all except the least able group. Five per cent or less mentioned 'ideal persons' from the following spheres: politics, science, the arts, history, travel, adventure, royalty.

There were interesting qualitative differences between boys and girls. The number of male 'ideal persons' named by girls was much larger than that of female

\* At the editor's request, levels of significance have been worked out although neither of the authors nor their statistical adviser consider that these provide much additional information. In a study of this nature a 'statistically significant difference does not necessarily mean a difference either of a practical significance or of scientific import' (McNemar, 1955). For this reason we simply report the level of significance reached and let the reader evaluate the finding.



'ideal persons' given by boys; for example, Jesus, other biblical characters and Shakespeare were frequently named by girls. The figures most frequently given by boys were: Nelson (7), Drake (4) and then footballers, e.g. Stanley Matthews, Billy Wright, etc. Girls named most frequently: Enid Blyton (10), Pat Smythe (8) and Florence Nightingale (6). Religious or moral leaders were mentioned more often by girls than boys and the same applied to royalty; on the other hand 15 boys but only one girl named war or travel heroes.

Thus 39 per cent of the boys' choices were associated with war, adventure, politics and sport, compared with 10 per cent among girls; on the other hand, only 5 per cent of boys but 19 per cent of girls chose either royalty or figures distinguished for moral, humanitarian or religious qualities. The least able group not only mentioned a comparatively higher proportion of 'ideal persons' from their own immediate circle, but the fields from which their 'remote' ideals were chosen was also rather more limited than that of any other subgroup.

### *Reasons for choosing the 'ideal person'*

Though every child was asked to give at least one reason for his choice, a few of the duller ones either wrote nothing or else repeated the name of their 'ideal person'; others, however, gave more than one reason. The reasons, grouped under a number of headings, are shown on Table 2. Duller children gave fewer reasons than abler ones. There was considerable unanimity in naming good, in the sense of being efficient, most frequently; good, in the sense of being kind, came next, except that adventurousness took second place for boys.

Some examples will illustrate the type and range of answers which were given. Mary, a bright girl, choosing Gladys Aylward as her 'ideal person', wrote: 'She was a very brave woman, a quality which I would love to possess! Her work among the Chinese shows that she was not afraid to take the good with the bad, the rough with the smooth, and she was not afraid of quelling a prison rebellion of dirty, rough men. The taking of 100 children over China to safety shows that she had great endurance.' Robert, a bright boy, wrote: 'I'd like to be Peter Scott because he has been all over the world from Equador to Thailand. Also he is very brilliant at drawing birds and other living creatures. He also has a part of the river Severn as his natural bird trust Head Quarters.' Peter, a dull boy, wrote: 'My Dad, because he is nice.'

### *2. The 'Moral Wickedness' Test*

The total and average number of responses given are shown on Table 3A. In addition to the answers listed in Table 3B, the following were mentioned but by less than 8 per cent of the children: not believing in God; A- & H-bomb; haughtiness; rudeness; blackmail; cheating; insulting behaviour; spitefulness; arson; smuggling; spying; spitting; making threats; disobeying; hateful behaviour; being cheeky; mental cruelty; treason; truanting; caning; slave trade; jealousy; drunkenness; strikes; divorce; smoking; throwing stones; forgery; witchcraft; playing cards on Sunday; not washing; boasting; ignorance; children being left alone at night; offences against old people; cutting up the Union Jack; being lost at 11 p.m.; phone tapping.



Table 1. 'Ideal person' according to sex and I.Q.

Category	Total N = 226		Boys N = 109		Girls N = 117		Able N = 77		Average N = 109		Low average and dull N = 40	
	N	%*	N	%	N	%	N	%	N	%	N	%
Immediate circle	57	25	25	23	32	27	10	13	29	27	18	45
Radio, T.V., films	49	22	21	19	28	24	19	25	21	19	9	23
Sport	34	15	23	21	11	9	17	22	15	13	2	5
Books—authors and characters	33	14	12	11	21	18	12	16	17	15	4	10
Religious/Moral	20	9	4	4	16	14	8	10	10	9	2	5
War figure/ 3 Ser- vices	10	4	10	9	—	—	4	5	6	6	—	—

\* Percentages are given to the nearest whole number in this and all subsequent tables.

Of the above differences those between the different ability groups choosing persons from their immediate circle were significant at the 1 per cent. level and for choosing sporting figures at the 5 per cent. level; for sport v. sex  $p$  was  $< 0.025$ .

Table 2. *Reasons for choosing 'ideal person' according to sex and I.Q.*

Reasons	Total N = 226		Boys N = 109		Girls N = 117		Able N = 77		Average N = 109		Low average and dull N = 40	
	N	%*	N	%	N	%	N	%	N	%	N	%
Good (efficient)	77	34	36	34	41	35	27	32	38	35	12	38
Good (kind)	45	20	15	14	30	25	13	15	22	20	10	31
Adventurous	25	11	21	20	4	3	9	11	14	13	2	6
Physical attributes	19	8	1	1	18	15	9	11	9	8	1	3
Powerful	18	8	8	7	10	8	7	8	8	7	3	9
Famous	14	6	7	6	7	6	6	7	6	5	2	6
Brave	11	5	9	8	2	2	5	6	5	4.5	1	3
Interesting people	7	3	3	3	4	3	3	4	3	3	1	3
Wealthy	4	2	2	2	2	2	4	5	—	—	—	—
Inventor/Scientist	4	2	3	3	1	1	2	2	2	2	—	—
Total No. of reasons	224		105		119		85		107		32	

\* Percentages refer to reasons and not to children.

Of the above differences the following were significant at the levels stated: adventurousness v. sex ( $p < 0.001$ ); physical attributes v. sex ( $p < 0.001$ ); good (kind) v. sex ( $p < 0.05$ ).



Table 3A. Total number and average number of 'Morally Wicked Actions'

	Total N = 226	Boys N = 109	Girls N = 117	Able N = 77	Average N = 109	Low average and dull N = 40
Total No.	744	317	427	311	334	99
Aver. No. per child	3.2	3	3.6	4	3	2.5

There was considerable unanimity not only on the six most 'wicked actions' but also on their respective degree of wickedness. However, the intellectually slowest group mentioned murder, stealing and lying less frequently and swearing more often than the other two ability groups. There were also some sex differences, a higher proportion of girls than boys considering stealing, cruelty to animals, lying and swearing to be wicked.

The most intelligent children mentioned not only a greater variety of types of moral wickedness—even under one heading such as murder—but also produced more subtle and abstract forms of 'wicked actions'. The following examples illustrate these two points. Susan offered these details for 'murder' alone: 'Chopping peoples heads off; hanging and burning alive; throwing people to the lions; burying alive; scalping someone; putting someone on the electric chair; blowing people up; suffocating or beating them to death.' John, a dull boy, wrote 'to kill a person; to swear; to bully somebody'. And Mary listed 'killing a young girl; forgery'. Among the dullest group there was a marked tendency to mix serious and trivial offences, for example: 'to murder'; 'shouting in the corridors'. Occasionally this also occurred among the average group. There were some rather individual replies such as David's, who wrote 'murder; starting the third world war; setting fire to corn and wheat; starting a native uprising; making people live in slums'. Lastly, some replies defined classification: making diesels instead of steam engines; the 'Blues' being relegated into division 2; leaving girls out of everything; forgetting the work of a housewife; leaving none of nature's beauty but building houses on it; teaching English all day; school dinners.

### 3. The 'Moral Incidents' Test

This test was devised to explore children's responses to hypothetical situations demanding decisions or judgments well within the experience of 11-year-olds. Each of the ten situations involved ethical or moral problems for which three alternative replies were presented. Children were allowed to work at their own pace. To facilitate identification for both sexes, the characters of four incidents were boys only or girls only respectively, while two incidents involved both. The 'forced-choice' type of design was chosen, partly because interpretation and comparison are more readily made than in open-ended tests, and partly in contrast to the two previous tasks which imposed no limitations. Two of the incidents (Nos. 1 and 5) are very similar to questions previously used by Piaget (1952).

For the purpose of discussing the results, the ten situations have been grouped together according to the four basic themes which were explored, namely honesty,

Table 3B. 'Morally Wicked Actions' according to sex and intelligence

Type of action	Total N = 226		Boys N = 109		Girls N = 117		Able N = 77		Average N = 109		Low average and dull N = 40	
	N	%*	N	%	N	%	N	%	N	%	N	%
Murder	193	85	96	87	97	82	71	92	94	85	28	70
Physical cruelty	143	63	70	64	73	62	54	70	63	57	26	65
Stealing	121	53	42	38	79	67	50	65	55	50	16	40
Cruelty to animals	68	30	24	22	44	37	30	39	25	23	13	33
Lying	41	18	13	12	28	24	18	23	22	20	1	3
Damage to property	30	13	14	13	16	14	15	19	10	9	5	13
Swearing	18	8	4	4	14	12	6	8	6	5	6	15
Blaming someone else	16	7	8	7	8	7	5	6	11	10	—	—
Kidnapping	14	6	8	7	6	5	11	14	3	3	—	—

\* Both the numbers and the percentages refer to the children who mentioned the particular deeds.

Of the above differences the following were significant at the levels stated: murder v. ability ( $p < 0.01$ ); stealing v. ability ( $p < 0.025$ ); stealing v. sex ( $p < 0.001$ ); cruelty to animals v. sex ( $p < 0.01$ ); lying v. sex ( $p < 0.025$ ); lying v. ability ( $p < 0.025$ ); swearing v. sex ( $p < 0.025$ ).



loyalty, responsibility and punishment; the order in which they were actually presented is indicated by the number assigned to each incident. The children were asked to underline whichever one of the three possible replies seemed in their opinion to be the most just or sensible way of meeting the situation in question. It was stressed that there were no right or wrong answers.

The four incidents, concerned with problems of honesty, read as follows:

*Incident No. 4*

Jill and Dora went into a sweet shop and found that the shop-keeper was out. There was some money lying on the counter. Jill took sixpence and Dora half-a-crown. They put the money in their purses. Do you think:

- (a) Jill should be blamed most?
- (b) Dora should be blamed most?
- (c) Both should be blamed the same?

*Incident No. 6*

George's family was so poor that they could not afford to buy sweets for his sister Pat who was ill in hospital. One day, George took three bars of chocolate from a shop and gave them to Pat in hospital. Do you think:

- (a) George should have taken the chocolates?
- (b) George should not have taken the chocolates?
- (c) George can be excused?

*Incident No. 7*

Fred and his sister Carol were on a 'bus one day and the conductor gave them a shilling too much change. Do you think they should have:

- (a) Told the conductor?
- (b) Kept the money and not told him?
- (c) Asked their mother what to do?

*Incident No. 8*

Jean was sent to get a dozen eggs for her mother. Instead of taking the eggs straight home she began to play ball with her friend. The ball hit the eggs and smashed them. Jean told her mother that she had slipped and fallen on the pavement. Do you think:

- (a) Jean should have told the truth?
- (b) Jean should not have told the truth?
- (c) Jean could be excused?

The replies received to these four incidents are shown in Table 4. In each case, the majority of children came down on the side of honesty and truthfulness, and there was little difference between boys and girls. In the two straightforward problems (Incidents Nos. 7 and 8) over 90 per cent of children gave the expected reply; but when the situations were made more complex—by the greater amount of money taken by one girl in incident No. 4 and by the extenuating circumstances of poverty and illness in incident No. 6—there was less unanimity.

Table 4. *Replies to incidents concerned with honesty according to sex and intelligence*

	Total N = 226		Boys N = 109		Girls N = 117		Able N = 77		Average N = 109		Low average and dull N = 40	
	N	%	N	%	N	%	N	%	N	%	N	%
<i>Incident 4</i>												
Jill blamed most	11	5.0	6	5.0	5	4.0	2	2.5	7	6.0	2	5.0
Dora blamed most	20	8.8	7	6.0	13	11.0	4	5.0	6	5.0	10	25.0
Both blamed same	195	86.2	96	89.0	99	85.0	71	92.5	96	89.0	28	70.0
<i>Incident 6</i>												
Should take chocolates	14	6.0	8	7.0	6	5.0	—	—	10	9.0	4	10.0
Should not take chocolates	145	64.5	66	60.5	79	68.0	56	73.0	63	59.0	26	65.0
Can be excused	67	29.5	35	32.5	32	27.0	21	27.0	36	32.0	10	25.0
<i>Incident 7</i>												
Tell conductor	214	95.0	102	93.5	112	96.0	77	100	103	94.5	34	85.0
Keep money	7	3.0	5	4.5	2	1.5	—	—	5	4.5	2	5.0
Ask mother	5	2.0	2	2.0	3	2.5	—	—	1	1.0	4	10.0
<i>Incident 8</i>												
Tell the truth	220	97.6	105	96.0	115	98.4	76	98.75	106	97.0	38	95.0
Not tell the truth	2	0.8	1	1.0	1	0.8	—	—	1	1.0	1	2.5
Can be excused	4	1.6	3	3.0	1	0.8	1	1.25	2	2.0	1	2.5

Of the above differences the following were significant at the levels stated:

Incident 4: ability ( $p < 0.005$ )

Incident 6: ability ( $p < 0.05$ )



The three incidents concerned with responsibility and loyalty, read as follows:

*Incident No. 2*

Joan is saving up to go on a school outing. She has been looking forward to this for a long time. But one Saturday her friend Mary asks her to go to the pictures. Joan can only go if she spends some of the money she has been saving for the outing. Do you think:

- (a) Joan should go to the pictures?
- (b) Joan should stay at home?
- (c) Joan should ask her mother what to do?

*Incident No. 3*

Jack was one of the best players in the school football team. One Friday evening the team had an important game to play, and needed all their best players to win. Jack went home to see a programme on television. He did not worry how the school did in the match. Do you think:

- (a) Jack should have played for the school?
- (b) Jack should have gone home to watch television?
- (c) Jack should have asked his mother what to do?

*Incident No. 10*

Molly has promised her friend to go to her house for tea. When the afternoon comes Molly finds she will miss her favourite programme on television. Do you think Molly should:

- (a) Go to tea with her friend?
- (b) Watch her television programme?
- (c) Ask her mother what to do?

The replies given to these three incidents are shown in Table 5. Greatest unanimity was shown over incident No. 3, while in incident No. 2 there was a marked division of opinion. Lastly, three incidents were concerned with problems of punishment and read as follows:

*Incident No. 1*

The teacher had warned Tom not to run wildly into the classroom. Tom forgot, ran in wildly and knocked a glass jar, full of water, on to the floor and broke it. Bill, walking into the room behind him, accidentally slipped and knocked twenty glass jars on to the floor and broke them. Should the teacher:

- (a) Punish Tom most?
- (b) Punish Bill most?
- (c) Punish both boys the same?

*Incident No. 5*

Stanley was a new boy in the class. In a test he saw other boys cheating, so he cheated too and copied many answers. Another boy, Peter, had cheated on a previous test and had been warned by the teacher not to do it again. But Peter cheated

Table 5. *Replies to incidents concerned with loyalty and responsibility according to sex and intelligence*

	Total N = 226		Boys N = 109		Girls N = 117		Able N = 77		Average N = 109		Low average and dull N = 40	
	N	%	N	%	N	%	N	%	N	%	N	%
<i>Incident 2</i>												
Go to pictures	4	1.6	1	1.0	3	2.5	1	1.25	1	1.0	2	5.0
Stay at home	92	41.0	41	37.5	51	43.5	37	47.75	43	39.0	12	30.0
Ask mother	130	57.4	67	61.5	63	54.0	39	51.0	64	60.0	26	65.0
<i>Incident 3</i>												
Play for school	212	94.0	101	92.5	111	95.0	77	100	104	95.5	31	77.5
Go home	—	—	—	—	—	—	—	—	—	—	—	—
Ask mother	14	6.0	8	7.5	6	5.0	—	—	5	4.5	9	22.5
<i>Incident 10</i>												
Go to tea	177	78.0	77	71.5	100	86.0	72	93.75	81	75.0	24	60.0
Watch T.V.	11	5.0	7	6.0	4	3.0	—	—	7	6.0	4	10.0
Ask mother	38	17.0	25	22.5	13	11.0	7	6.25	21	19.0	12	30.0

Of the above differences the following were significant at the levels stated:

Incident 3: ability ( $p < 0.001$ )

Incident 10: ability ( $p < 0.001$ )

Incident 10: sex ( $p < 0.01$ )



Table 6. Replies to incidents concerned with punishment according to sex and intelligence

	Total N = 226		Boys N = 109		Girls N = 117		Able N = 77		Average N = 109		Low average and dull N = 40	
	N	%	N	%	N	%	N	%	N	%	N	%
<i>Incident 1</i>												
Punish Tom most	114	50.5	57	52.5	57	48.5	56	73.0	53	49.0	5	12.5
Punish Bill most	22	9.5	15	14.0	7	6.0	2	2.5	10	9.0	10	25.0
Punish both same	90	40.0	37	33.5	53	45.5	19	24.5	46	42.0	25	62.5
<i>Incident 5</i>												
Punish Stanley most	16	7.0	11	10.0	5	4.0	3	3.75	7	6.0	6	15.0
Punish Peter most	75	33.0	42	38.5	33	28.0	34	44.0	34	31.0	7	17.5
Punish both boys same	135	60.0	56	51.5	79	68.0	40	52.25	68	63.0	27	67.5
<i>Incident 9</i>												
Be warned	25	11.0	13	12.0	12	10.0	2	2.5	15	13.5	8	20.0
Make up time after school	123	54.5	60	55.0	63	54.0	35	45.25	62	58.0	26	65.0
Be punished	78	34.5	36	33.0	42	36.0	40	52.25	32	28.5	6	15.0

Of the above differences the following were significant at the levels stated:

- Incident 1 : ability ( $p < 0.001$ )  
 Incident 5 : sex ( $p < 0.05$ )  
 Incident 5 : ability ( $p < 0.025$ )  
 Incident 9 : ability ( $p < 0.001$ )

again and the teacher caught him when he had copied one answer. Then the teacher also caught Stanley after he had copied many answers. Do you think the teacher should:

- (a) Punish Stanley most?
- (b) Punish Peter most?
- (c) Punish both boys the same?

#### *Incident No. 9*

In a school, there was a strict rule that children should always be on time. Dick came to school late after having been already reminded twice about this rule. Do you think Dick should:

- (a) Be warned and let off again?
- (b) Asked to make the time up after school?
- (c) Be punished?

The replies given to these three incidents are shown on Table 6. Overall there was much less unanimity on this group of problems than on the two previous ones.

### IV. DISCUSSIONS AND CONCLUSIONS

Towards the end of the junior school stage a majority of children seem to choose as their ideal figure characters personally unknown to them. Radio, T.V. and film personalities predominated. Bray's (1962) findings are very similar to ours since among his sample of 2415 10- to 11-year-olds 83 per cent of boys and 79 per cent of girls chose 'remote' admired persons.

The younger the child or the less able, the more likely it is that his 'ideal person' will be chosen from his immediate circle of relatives and personal friends. In a similar study with boys in a secondary modern school (Edwards, 1959) an even smaller proportion named 'ideals' who were personally known to them, but instead chose personalities from the entertainment or sports world. In a much earlier investigation (Macauley & Watkins, 1926) it was found that people from the child's immediate circle were most frequently named as 'ideals', but the second and third place were taken by characters from fiction and adventure stories; this probably reflects the fact that in those days there were few radio and no television sets to create popular personalities. Royalty is also mentioned far less frequently by the modern child.

With regard to sex differences, boys on the whole chose male 'ideal persons' whereas girls' ideals were often of the opposite sex. Boys mentioned a relatively greater proportion of characters associated with war, adventure, politics and sport whereas girls chose a higher proportion of figures distinguished because of moral or religious qualities or for being of royal blood. Again Bray's (1962) findings with children of the same age, were closely similar in this respect.

Compared with the secondary modern schoolboys, moral or religious leaders were more often chosen by our somewhat younger children, especially girls; none of them named themselves, whereas quite a few of the 14-year-olds did so, giving as the reason 'I am happy as I am'. However, the most frequently mentioned reason



for the choice of the ideal person was the same for the younger and the older children, namely 'good', in the sense of efficient. But whereas the younger boys gave 'adventurousness' and 'good' in the sense of kind, as their second and third reason, the older boys cited 'fame' and then 'wealth'.

The more intelligent children in both the primary and secondary schools gave, on the average, more reasons for their choice. The differences found between the sexes conform to what one would expect to result from the sex-typing prevalent in our society; adventurousness and braveness are qualities valued by boys, while goodness, in the sense of kindness, and physical attributes are of greater importance in the eyes of girls. The fact that goodness, in the sense of efficient, is valued by girls as much as it is by boys, may be symptomatic of how far the emancipation of women has succeeded by now. One wonders why twice as many of the slower children should mention good, in the sense of kind, compared with the able group: is it because the former are in greater need of and therefore more appreciative of this quality? It is interesting to note that despite their growing up in an affluent and acquisitive society, wealth and riches were given such very low priority. This was also the case in Bray's (1962) study. Perhaps it is mainly during adolescence that the idealism of the earlier stage becomes contaminated by the prevailing climate of opinion and values.

As one would have expected from previous studies, the 'wicked deeds' mentioned most frequently were murder, physical cruelty, stealing and cruelty to animals. This agrees closely also with the actions named by the secondary modern school boys (Edwards, 1959). However, whereas the younger children listed next lying and damage to property, the older boys named blackmail, the H-bomb and sabotage; they also mentioned corporal punishment more frequently. The more intelligent children, both among the younger and older samples, tended to produce longer and more varied lists of 'wicked actions', including more subtle and abstract kinds; they also frequently gave a greater variety of any one form of 'wickedness'. Less able children usually mentioned only one or two items and often placed serious and trivial forms of wickedness in juxtaposition. It is interesting to note the very low order of priority accorded to disobedience and cheekiness; on the other hand, one wonders whether the high position given to 'cruelty to animals' would be found in any but the Anglo-Saxon culture. It is also probably a reflection of the moral climate of our society that murder and cruelty to animals are given pride of place while the A- and H-bombs and mental cruelty are but rarely mentioned; yet the former constitutes a far less likely threat to the welfare of the individual, or humanity at large, than the latter.

The findings from the 'Moral Incidents' Test have a number of suggestive implications regarding the understanding of moral concepts by children of this age. To straightforward issues, such as telling the truth when an accident has led to some damage (incident No. 8) or when too much change is given in error (incident No. 7), the 'right' solution or behaviour was known to practically everyone. (Of course, this does not necessarily mean that a similarly high proportion would in fact take the 'right' action.) When the issue was more subtle or complex, the more intelligent children could still discriminate between right and wrong behaviour whereas a



proportion of the slower ones apportioned blame according to the amount of money stolen. Extenuating circumstances led over a quarter of all the children to condone theft, while a few of the slower ones actually approved of it. Thus it seems that the duller child's moral judgment is prone to become confused or blurred in the face of complexity.

Loyalty to school and to one's team seemed to be a generally accepted concept; some uncertainty was shown only by a small proportion of the dull children, some 20 per cent of whom felt that advice from mother should be sought (incident No. 3). With regard to keeping a promise of a personal and rather social kind, the trend of replies was similar but among all subgroups a small number were in favour of asking mother's advice, the proportion being highest for boys and the dumbest group where it reached about 25 per cent (incident No. 10). Thus it would appear that children of this age appreciate that all promises should be kept but that breaking certain kinds of commitments may be morally less reprehensible than others. At first sight, it is rather surprising that more than half of the children should think mother's advice is needed in what seems neither a complex nor a very unusual situation (incident No. 2). It may be due to the fact that it presents a conflict of loyalty (to school and to a friend) as well as the dilemma of choosing between two pleasures, one immediate and one long-deferred. Alternatively, or possibly in addition, it lies in mother's power to resolve the difficulty by offering to pay for the cinema ticket and thus asking for her view may be considered quite a constructive alternative. It is worth noting that very similar proportions of boys and girls chose this reply and that the difference between the brightest and dumbest group was also comparatively small.

The results obtained from the three incidents dealing with punishment are perhaps the most interesting. That children often tend to be less lenient than adults, when a rule is broken, receives some confirmation from the fact that only a very small minority favoured a third warning (incident No. 9). While there was no sex difference, the least able children were most inclined to give the offender another chance. This may be prompted by their unconscious awareness that they themselves are in need of leniency because of their tendency towards impulsiveness and forgetfulness. 'To make the punishment fit the crime' was the preferred course of action of over half of the children, except the most able group, who stood out as being the most punitive among all the subgroups. May this be a reflection of the high demands and pressures which adults exert upon able children and which they in turn tend to introject?

The aim of the last two incidents (Nos. 1 and 5) was to explore to what extent children's views on punishment are influenced by the intent or motive of the culprit and by the actual size of the damage or offence. It would seem that the great majority of children took into consideration extenuating circumstances (lack of intent in incident No. 1 and lack of knowledge in incident No. 5). The small minority who considered that the 'less guilty' offender should be punished most was probably influenced by the fact that he committed the 'bigger' offence (i.e. broke more jars or copied a greater number of items). Yet despite the general recognition of 'greater culpability', the number of children who considered that the same punishment should be meted out to both offenders was considerable. The most marked differ-



ence in viewpoint was found between the different ability groups: the most intelligent children gave less weight to the actual consequences of the action and more to the motive or intention underlying it; the least able group tended to pay greater attention to the amount of damage or cheating and a higher proportion than among the other two groups favoured what might be considered an easy way out, namely meting out the same punishment to both offenders.

What, then, are the practical and theoretical implications of this study? That younger and less able children tend to choose their 'ideal person' from among personally known friends and relatives would be predicted from the fact that learning as well as identification proceed from the known to the unknown; it requires more mature imaginative power to project oneself into the role of someone remote from first-hand experience.

Now that women are increasingly being given equality, both educationally and vocationally, one wonders why many girls should still be choosing their ideals from among the opposite sex. Is it because there are fewer outstanding women than men? Or is there in our teaching—whether it be directly in history lessons in school, informally at home or through the mass media of information and entertainment—a conscious or unconscious bias against women? Or is it that despite the equalizing of opportunities open to men and women, traditional patterns of sex typing continue to be used so that girls still tend to consider the possession of outstanding personal qualities a predominantly male prerogative?

It is not perhaps unexpected that in a society as highly competitive as ours 'good', in the sense of being efficient at doing something, should be the most frequent reason for choosing an 'ideal person'. And it is worth noting that the least able group mentioned this quite as frequently as the abler children. Studies of child development usually emphasize the matter-of-fact realism of the junior school child in contrast with the idealism of adolescents. Our results suggest that present-day materialism may have a greater influence on older children, since among them 'fame' and 'wealth' had transplanted 'good', in the sense of kind, as admired qualities.

The evidence we obtained from the 'moral incidents' confirm the work of Harrower (1934), Lerner (1937), MacRae (1950, 1954) and Morris (1958) in modifying and amplifying Piaget's conceptual framework. Intellectual ability does seem to play a part in moral understanding, less able children tending to attach more importance to the actual consequences of actions rather than to intent or culpability. More advanced stages of moral development are reached earlier by more intelligent children. By the age of 10 to 11 years, then, many but not all children have attained a considerable degree of moral autonomy, while slower pupils retain modes of thinking which Piaget would describe as belonging to the rigidity of the transcendental stage. However, there were also many individual differences—another finding which workers, repeating Piaget's experiments, have noted. It is hoped to investigate later whether parental socio-economic status and different teaching methods influence the moral judgment of the children in our sample.

In summary, our results confirm the view (Burt, 1954; Tansley & Gulliford, 1960) that the moral concepts and judgments of duller children tend to be less mature, less clear and more limited than those of their brighter contemporaries. The latter



showed, at least on pencil and paper tests, socially more acceptable views, a more subtle understanding of moral issues and less dependency on adult advice.

Lip service is very widely paid to the importance of moral education, particularly in a rapidly changing society. Yet in practice we give insufficient recognition or time to the development of children's moral understanding or concepts. What is needed is a 'blueprint for an educational system which would encourage the growth of an integrated personality. Such a system would by-pass the split between intellectual and moral education. Such a study would be difficult. But if it were successful it could lead to a new conception of education' (Huxley, 1962). 'The moral education of young people needs to be tackled in a planned and concerted way by teachers, clergy, doctors. . . . We have not really begun to work systematically in this field' (Jeffreys, 1962). Among the many reasons for this are likely to be the current emphasis on objectivity, the lack of, almost fear of, idealism and the widespread pre-occupation with examinations together with the search for measurable or testable results. Added to this there is a dearth of research in this country into children's moral development and understanding. Before we do more we need to know more. Though the field is difficult and complex, we neglect its systematic study at a high cost in terms of wasted ability, maladjustment and delinquency.

#### REFERENCES

- ALLPORT, G. W. (1937). *Personality*. Holt: New York.
- BRAY, D. H. (1962). A Study of Children's Writing on an Admired Person, *Educ. Rev.* **15**, 44-53.
- BURT, Sir C. (1954). *The Causes and Treatment of Backwardness*. London: Univ. Press.
- EDWARDS, J. B. (1959). Certain Moral Attitudes among Boys in a Secondary Modern School. Unpublished M.A. Thesis. Univ. of Birmingham.
- EYSENCK, H. J. (1953). *The Structure of Human Personality*. London: Methuen.
- EYSENCK, H. J. (1960). The contribution of learning theory. *Brit. J. educ. Psychol.* **30**, 11-21.
- HARROWER, M. R. (1934). Social status and the moral development of the child. *Brit. J. educ. Psychol.* **4**, 75-95.
- HARTSHORNE, H. & MAY, M. A. (1930). *Studies in the Nature of Character*, 3 vols. New York: Macmillan.
- HAVIGHURST, R. J. & TABA, H. (1949). *Adolescent Character and Personality*. New York: Wiley.
- HAVIGHURST, R. J. & NEUGARTEN, B. L. (1955). *American Indian and White Children*. Chicago: University Press.
- HEMMING, J. (1957). Some aspects of moral development in a changing society. *Brit. J. educ. Psychol.* **27**, 77-88.
- HUXLEY, Sir JULIAN (1962). Fawley Foundation Lecture, reported in *The Times Educ. Supplement*, Nov. 23rd.
- JEFFREYS, M. V. C. (1962). *Personal Values in the Modern World*. Pelican Original.
- JONES, V. (1954). Character development in children—an objective approach. In *Manual of Child Psychology* (ed. L. Carmichael). London: Chapman & Hall.
- LENER, E. (1937). *Constraint Areas and Moral Judgement of Children*. Wisconsin: Menasha.
- MACAULAY, E. & WATKINS, S. H. (1926). An investigation into the development of the moral conceptions of children. Parts 1 and 2, *The Forum of Education*, IV.
- MACRAE, D. Junr. (1950). The Development of Moral Judgement in Children. Ph.D. thesis, Harvard University.
- MACRAE, D. Junr. (1954). A test of Piaget's theories of moral development. *J. abnorm. (soc.) Psychol.* **49**, 14-18.



- McGRATH, M. C. (1923). A study of the moral development of children. *Psychol. Monogr.*, No. 44.
- McNEMAR, Q. (1955). *Psychological Statistics*. London: Chapman & Hall.
- MALLER, J. B. (1934). General and specific factors in character. *J. Soc. Psychol.* 5, 97-102.
- MORRIS, J. F. (1958). The development of adolescent value-judgements. *Brit. J. educ. Psychol.* 28, 1-14.
- PETERS, R. S. (1960). Freud's theory of moral development in relation to that of Piaget, *Brit. J. educ. Psychol.* 30, 250-258.
- PIAGET, J. (1932). *The Moral Judgement of the Child*. London: Kegan Paul, Trench and Trubner.
- SYMONDS, P. M. (1931). *Diagnosing Personality and Conduct*. New York: D. Appleton—Century.
- TANSLEY, A. E. & GULLIFORD, R. (1960). *The Education of Slow-learning Children*. London: Routledge and Kegan Paul.

*Original manuscript received 11 January 1963*

*Revised manuscript received 29 May 1963*

## Acquiescence—Measurement and Theory\*

By JOHN MARTIN

*University of Western Australia*

A review of the methods of measuring acquiescence suggests that there is insufficient justification for regarding acquiescence as a single trait. As this assumption is the basis for the construction of independent general measures of acquiescence, such scales call for investigation. Seven measures of acquiescence were administered to 85 male householders of heterogeneous social background. Included were three scales presumed to be independent general measures of the tendency to acquiesce, two sets of acquiescence scores derived from different social desirability scales, and two derived from different authoritarian scales. Two factors were found, the first being defined by acquiescence from scales employing plausible generalities (particularly 'Rational Authoritarianism') and the second by Welsh's R-Scale. Both factors were shown to be independent of social desirability, but all other 'independent' measures of acquiescence were shown to be confounded with social desirability or authoritarianism or both.

It is suggested that the first factor is generated by the subjects' tendency to acquiesce in what he perceives to be generally accepted as rational or authentic, this explanation being a parallel to that advanced for the origin of social desirability. A three-factor theory of response style is put forward to account for the relationships found.

### INTRODUCTION

Acquiescence is customarily defined as a tendency to reply 'Yes', 'True', or 'Agree', to questionnaire items regardless of their content. It is regarded as a single tendency which interferes with the response to the content of the items so that subjects who are prone to acquiescence tend to score highly on all positively keyed tests undeterred by different—or even contradictory—contents. If it can be shown that the evocation of acquiescence bears a lawful relation to content, so that it is evoked discriminatively by different types of item, then this definition and the notions of how acquiescence operates, may have to be revised.

Two methods of measuring acquiescence have been evolved. One, historically prior, is based on the use of 'reversed' scales in which acquiescence is recognizable as a tendency leading to self-contradiction. The other method uses a large sample of items and relies on scoring the number of positive responses given without regard for the content or form of the items except to require that they be heterogeneous.

The second method rests directly on an assumption that acquiescence is a single tendency which is general over tests; the first makes no such assumption. The contrast is best seen if the two methods are examined in some detail.

Figure 1, which is essentially an empty scattergram for a 10-item positive and 10-item negative scale is a convenient way to illustrate the relationships between

\* This study is part of an investigation for the Ph.D. degree at the University of Western Australia under the supervision of Professor K. F. Walker. It has been made possible by research grants from the University of Western Australia, Key West Ltd., and the Carnegie Corporation of New York, to whom thanks are due. None of the bodies supplying funds is responsible for any of the views expressed in this report.



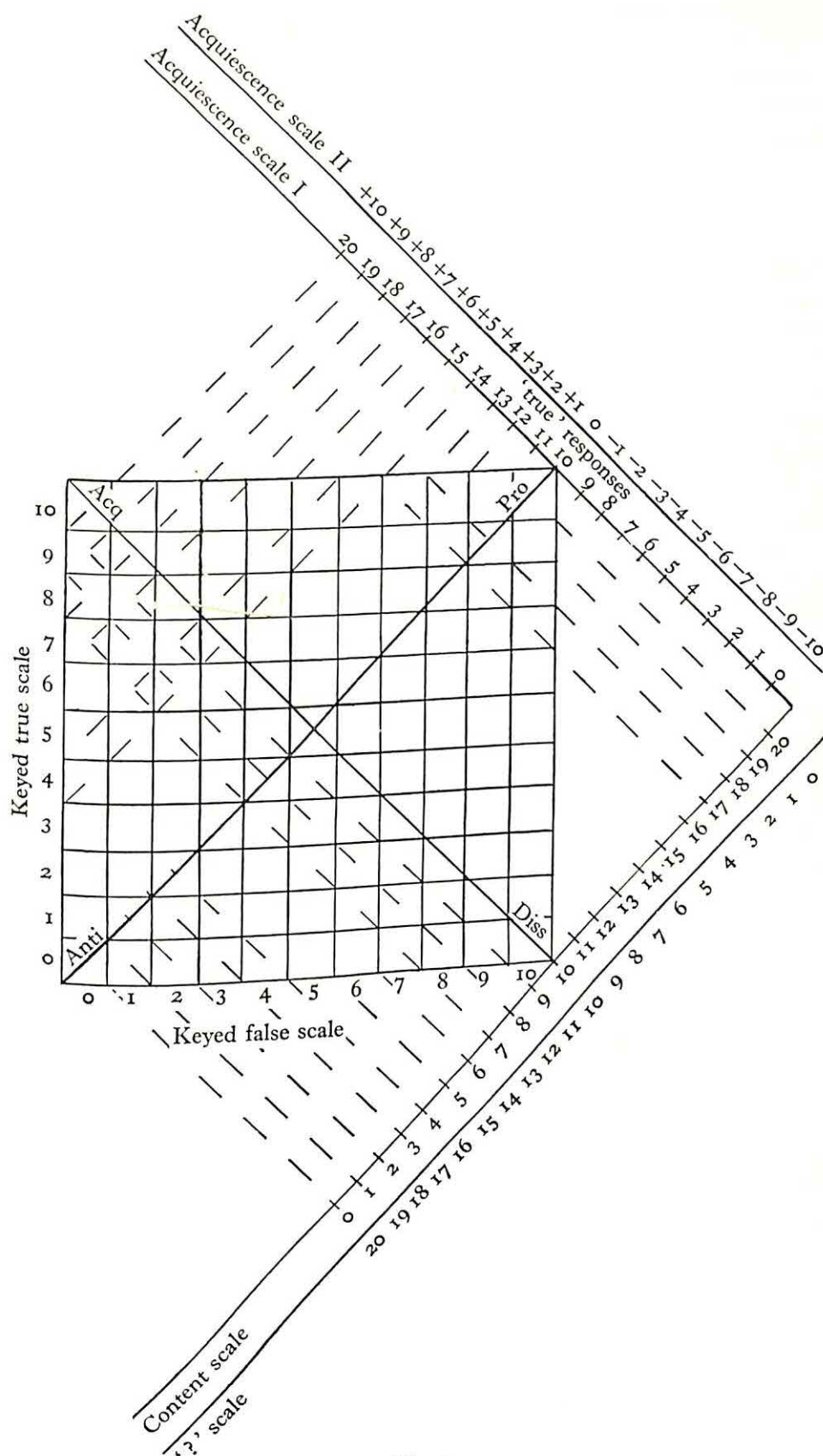


Fig. 1.

positive and negative halves of the scale and assists in appreciating how the correlation between the two versions reveals the presence or absence of acquiescence. This correlation may vary from  $+1.0$  (complete absence of acquiescence) to  $-1.0$  (when content is completely ineffective in determining replies). In the former case the scores when plotted will be along the diagonal labelled 'Anti-Pro' and in the latter along the diagonal 'Acq-Diss'. Deviation from these distributions indicates some admixture of acquiescence and content in the responses. Various scoring conventions for both acquiescence and content are indicated on the scales parallel to the diagonal but outside the square in Fig. 1. Three further general points are obvious from this diagram: (a) The reference axes on which content and acquiescence are measured are always orthogonal regardless of whether the positive and negative scales contain an equal number of items, (b) despite this, content and acquiescence are free to correlate (within limits) in any empirical case, (c) examination of the plotted scores will reveal more about the relations between content and acquiescence variance than will computation of  $r$  between positive and negative halves of the scale.

It should also be noted that where a 'Don't know' or '?' response option is offered, the number of '?' responses sets an upper but not a lower limit on the content score and has a complex relationship with both upper and lower limits of the acquiescence score. Because of these experimentally dependent relationships it would seem best to follow Cronbach's (1946) suggestion to eliminate the '?' response.

Similarly, it may be seen from Fig. 1 that, although the scales for content and acquiescence are set orthogonal, they are not independent in the sense that any score on one may be associated with any score on the other. Extreme scores on one imply middle scores on the other, and, in general, the score on one sets upper and lower limits to the score on the other. In such a case the variables may be described as having a reciprocal moderating relationship. This is a purely mechanical function of their limited interdependence and has no psychological significance, but it should be noted that it is logically consistent with an interpretation of acquiescence as being due to ambiguity generated in the subject-test interaction, because in so far as responses are governed by the content, ambiguity is absent, and vice versa. This would accord both with Cronbach's (1946) suggestion and with Banta's conclusions of 1961.

Because of the difficulties associated with the experimentally dependent measures produced by the reversed scale model, an independent scale of acquiescence has recommended itself to some researchers. This brings us to the second class of measures of acquiescence. All scales of this type are based, perforce, on the assumption that acquiescence is a general and unitary tendency which may be elicited from the subject by any set of items which is sufficiently heterogeneous. Examples of this type of scale are Bass' (1956) Social Acquiescence (or 'Babbitry') and Couch & Keniston's Agreement Response Scale (1960). Welsh's R-Scale has also been identified as acquiescence by Jackson & Messick (1961). Constructors of this type of scale have been content to support the assumption of generality by more or less significant correlations found between their scale and the positive



and negative halves of various personality measures, and the question of singleness is usually considered to be settled by definition.

Two important points should be noted about this method. In contrast to the first method discussed (which allows each measure of acquiescence to be specific to the scale from which it is derived) all specificity is here suppressed in the interests of the assumption of singleness. Although the operational definition of acquiescence involved is perfectly sound, it is not necessarily the same as that arrived at by the previous method, which is equally sound in its own right.

Secondly, by taking the largest practicable sample of items and retaining those which correlate best with the total 'Agree' or 'Yes' responses, the method increases the probability that if there exist clusters of items which function in this way because of their content or structure, such items will be included in the final measure. Two such clusters are known in the literature. One is concerned with social desirability and one with authoritarian attitudes. Should an acquiescence scale inadvertently contain items which align themselves with such clusters, then it will correlate with the relevant source scales as an artifact. This would be an embarrassment both in discussion of the part played by acquiescence in a scale, and in testing for evidence of a substantive personality trait of acquiescence.

Direct evidence of the unidimensionality and generality of acquiescence across tests is scarce, and where it has been sought the results do not favour the assumption of a single, non-specific tendency (McGee, 1962).

This study investigates the relationships among seven measures of acquiescence, three being independent measures, two being derived from 'reversed' scales of social desirability and two from 'reversed' authoritarian scales.

#### METHOD

Seven questionnaire measures of acquiescence were administered to a sample of 100 male householders in Western Australia; 85 usable protocols were returned. The sample was heterogeneous as regards age, education, job level and religious and political affiliation. The first three of these sociological variables showed no extreme departures from normality.

The measures of Acquiescence included in the battery were:

1. Bass Social Acquiescence. Fifty-six items taken from Bass (1956). All scored 'Agree'.
2. Agreement Response Set. Fifteen items drawn from Couch & Keniston (1960) constituting the short form of their 'Overall Agreement Scale'. Seven response options are offered for each item with weights from 1 (strongly disagree) to 7 (strongly agree).
3. Welsh's R-Scale. Although constructed by Welsh (1956) as a measure of repression, this scale was found by Jackson & Messick (1962) to identify the factor which distinguished subscales scored 'True' from those scored 'False' on the MMPI. It is therefore regarded by these authors as a measure of acquiescence. In the Jackson & Messick study the R-factor was found to be orthogonal to social desirability. This relationship moderating relationship appeared between R and social desirability. The R-scale also fell on is not due to score restrictions as in the case discussed above. Wiggins (1962).
4. Acquiescence drawn from the Marlowe-Crowne Scale (1960) of Social Desirability. This scale consists of 33 items of which 18 are scored 'True' and 15 'False'. A graph of the correlation surface for the two halves of the scale indicated the presence of considerable acquiescence variance. The scale was therefore scored for acquiescence by counting all items checked 'True'.

5. Two other social desirability scales, Edwards' (1957) and Jackson & Messick's Dyl (1962) were used. The 29 items drawn from Edwards' were all scored 'False' and the 48 items of the Dyl were all scored 'True'. The correlation between these positive and negative forms was high ( $r = +0.614$ ,  $p < 0.005$ ) and acquiescence variance was therefore comparatively low. Nevertheless, the two scales were combined to give a measure of the small acquiescence generated.
6. Rudin's Authoritarianism Scale. This consists of 19 items, six scored 'True' and the remainder 'False'. The scale was developed by Rudin (1961) to measure 'Rational Authoritarianism' as opposed to the 'Irrational Authoritarianism' of the more familiar California F-scale. Rudin reported that when scored for content this scale correlated  $+0.04$  with the F-scale. In the present study the Rudin scale was scored for acquiescence by counting all items checked 'True'.
7. Two short forms of the F-scale, one positive and one negative, were included in order to obtain a measure of the much studied acquiescence from this scale. These were the five-item version used by Srole (1956) and five negative items written by Couch & Keniston (1960). These were also scored for acquiescence.

### RESULTS

Intercorrelations among the seven measures of Acquiescence are given in Table 1. Nine of these correlations are significant at or beyond the 0.5 per cent level, the largest appearing between the acquiescence drawn from the Rudin Scale and that drawn from the short form of the F-scale.\* Bass' Social Acquiescence is also closely related to the acquiescence from the Rudin Scale. Welsh's R-Scale, on the other hand, has negative correlations with all other scales, most of them low, but the correlation with acquiescence generated by the Marlowe-Crowne Social Desirability Scale is highly significant. Inspection of this matrix suggests that at least two factors are present.

Table 1

	1	2	3	4	5	6
1						
2	361†					
3	-108	-285*				
4	298†	213	-367†			
5	479†	370†	-171	365†		
6	501†	328†	-081	168		
7	252*	215*	-078	162	439†	
					247*	517†

\* Significant at the 5 per cent level.

† Significant at the 1 per cent level.

‡ Significant at the 0.5 per cent level.

These data were factor-analysed by the principal components method in the computer laboratory of the University of Western Australia and yielded the seven factors shown in Table 2. The first two of these have latent roots greater than 1.0 and must therefore be retained according to Kaiser's (1960) criterion. The latent roots were plotted and it was found that III, V, VI and VII could be fitted to a straight line and may therefore be discarded. The latent root for Factor IV

\* Notwithstanding the high correlation between their acquiescences, the Rudin Rational Authoritarian Scale and the short (positive) form of the F-scale correlated only 0.117 (n.s.) when both scales were scored for content. This supports Rudin's claim that the contents of the Rational and the Californian Authoritarian scales are independent.



departed slightly from this line and may contain some systematic variance, possibly due to the operation of a different response style (e.g., Extreme Set) in the Couch and Keniston Scale. The amount of variance accounted for was not considered large enough to warrant retention of this factor.

The first two factors were rotated graphically to the structure shown in Table 3.

Table 2. *Unrotated factors*

		I	II	III	IV	V	VI	VII
Bass' Social Acquiescence	1	0.73	0.15	-0.39	0.05	0.40	-0.27	0.25
Agreement Response Set	2	0.63	-0.15	-0.08	-0.66	-0.29	-0.22	0.05
Welsh's R-Scale	3	-0.39	0.72	-0.38	0.17	-0.33	-0.21	0.06
Marlowe-Crowne SD (Acq.-from)	4	0.55	-0.53	0.01	0.52	-0.21	-0.28	0.16
Dyl + Ed. SD (Acq.-from)	5	0.74	-0.01	-0.35	0.13	-0.24	0.50	0.12
Rudin's Authoritarianism (Acq.-from)	6	0.74	0.44	0.13	-0.00	0.21	0.10	0.44
F-scale short form (Acq.-from)	7	0.56	0.42	0.62	0.17	-0.17	0.06	0.27
Latent roots		2.78	1.21	0.83	0.77	0.53	0.50	0.37

This solution capitalizes on the fact that Welsh's R-scale has been shown by Jackson & Messick to mark a factor of acquiescence orthogonal to social desirability. The R-scale is used to define Factor II on the assumption that any solution which fails to place a factor through this scale runs the risk of arriving at factors which correlate with social desirability. Acquiescence from the F-scale and Bass's

Table 3. *Rotated factors*

	I	II
1 Bass' Social Acquiescence	0.71	-0.21
2 Agreement Response Set	0.46	-0.47
3 Welsh's R-Scale	0.00	0.81
4 Marlowe-Crowne SD (Acq. from)	0.22	-0.72
5 Dyl + Ed. SD (Acq. from)	0.64	-0.36
6 Rudin's Authoritarianism (Acq. from)	0.85	0.04
7 F-scale short form (Acq. from)	0.69	0.11
	2.3723	1.5828
Sums of squares	34	23
Per cent total variance	59.98	40.02
Per cent extracted variance		

Social Acquiescence both have high loadings on Factor I but the acquiescence from the Rudin Scale offers the most definitive measure of this factor. The distribution of tests in the factor space is similar to (but not so complete as) the circular array from which Jackson and Messick deduced a reciprocal moderating relationship between social desirability and the R-scale. It would therefore seem likely that a similar relationship exists between the R-scale and acquiescence drawn from the authoritarian scales. This might suggest that the type of acquiescence induced by the authoritarian scales is closely related to social desirability. That this is not the case is indicated by the fact that when our social desirability measures are

scored for social desirability (content) their relations with the scales marking Factors I and II are insignificant in every case (Table 4). The two factors of acquiescence must therefore be regarded as largely independent of social desirability.

Table 4

	Rud. Acq.	F. Acq.	R
Social desirability	132(n.s.)	015(n.s.)	-063(n.s.)
Marlowe-Crowne SD	042(n.s.)	014(n.s.)	069(n.s.)

#### INDEPENDENT MEASURES OF ACQUIESCENCE AND OTHER RESPONSE STYLES

The hypothesis that generalized scales of acquiescence are confounded with social desirability is confirmed by the correlations shown in Table 5. Both Bass' Scale and the Couch & Keniston Scale show highly significant relationships with one or other of the social desirability measures. Welsh's R-Scale, in contrast, proves to be relatively independent of social desirability, as in the Jackson & Messick study. Recent claims that social desirability is chiefly acquiescence therefore probably arise because the measures of acquiescence used are themselves confounded with social desirability.

Table 5

	Ed.	DY 1	Bal. SD
BSA	158	480*	321*
R	+122	-210	-063
ARS	-346*	067	-167

\* Significant at the 0.5 per cent level.

Evidence concerning the hypothesis that the general measures of acquiescence will be confounded also with authoritarian item clusters is to be found in Table 6, where it is seen that the Bass Scale in particular is significantly related to both authoritarian scales. The R-scale & the Couch & Keniston Scale are related to the positive form of the F-scale but at a lower level. Hence this hypothesis is also confirmed. However, the R-scale and the Rudin Scale are independent whether the Rudin is scored for content or acquiescence.

Table 6

	Rudin Scale (Content)	F-Scale (Content)
Bass Social Acquiescence	312†	510†
Welsh's R-Scale	085	-240*
Agreement Response Scale	-141	244*

\* Significant at the 5 per cent level.

† Significant at the 0.5 per cent level.

#### DISCUSSION

The demonstration that two independent factors are necessary to account for the acquiescence drawn from a number of measures of this variable shows that the unidimensionality of acquiescence cannot be assumed. Hence the measurement problem is not to be solved by the construction of a single independent scale. The



data suggest that the measures required may be limited to two, and that these are capable of development from the Rudin Authoritarian Scale and from Welsh's R-Scale respectively.

To assert two factors of acquiescence is to assert a paradox, for by definition acquiescence is independent of content—all content—and to be a satisfactory concept should appear as a unitary tendency to agree in any situation. The results found here are to be explained only if different contents elicit acquiescence independently, for the existence of two orthogonal factors of acquiescence implies that the subject who acquiesces in one situation may or may not acquiesce in the other. Messick (1962) also found two orthogonal factors of acquiescence, but was unable to explain the second factor unequivocally and suggested that it represented a form of acquiescence interpretable in terms of content relationships. Fortunately the scales marking the two factors found in the present study show differences which lead to a more definite statement of possibilities.

All the scales which define Factor I (Bass' Social Acquiescence, Rudin Authoritarianism and the F-scale) consist of *generalities or aphorisms*, sometimes concerned with authoritarianism but not always so. The scales defining Factor II (Welsh's R-Scale, Marlowe and Crowne's SD Scale) contain no such general statements but consist of personal questions about the subject. As early as 1955, Leavitt, Hax & Roche concluded that acquiescence from the F-scale was connected with a tendency to accept plausible generalities, and Factor I of this study supports their suggestion.

Welsh's R-Scale has maintained its independence in this study as in Jackson's & Messick's 1962 factor analysis. It is heterogeneous in content but quite distinct in the form of its questions from the Factor I scales, being composed of typical self-descriptive items from the MMPI. Welsh described it as repression, and Jackson & Messick found it gave the best measure of the tendency to reply 'False' when SD was neutral. These two interpretations are not necessarily conflicting. In finding R to be orthogonal to the acquiescence from the authoritarian-type scale this study adds to the picture presented by Jackson & Messick suggesting that 'pure' acquiescence as represented by the R-scale will appear only when both the influence of plausible generalities and SD are neutral.

The R-scale therefore retains the status given it by Jackson and Messick as the best measure of 'pure' acquiescence but is seen to be inadequate as a 'general' measure of acquiescence. For instance, questionnaires from which R was partialled out could still retain acquiescence of the authoritarian (Factor I) type. Similarly, authoritarian acquiescence is not an adequate 'general' measure. For this reason studies such as Eysenck's (1962) investigation of the role of acquiescence in the N- and E-scales need to use both measures before they can be considered conclusive.

### Response style theory

The above results could be explained by the following theory:

The above results could be explained by the following theory: When a subject is confronted with a questionnaire item the form or content of the item may evoke a tendency to respond in terms of social desirability. In this case acquiescence (of any type) will be minimal and responses will tend to be consistent as between an item and its contradictory. If, however, such clear guidance is not available because the item is



neutral for social desirability (and he may be idiosyncratic in finding it so) the subject may be guided in his reply by what he considers to be rational or authentic. Inevitably such a judgment would be based, like social desirability, on notions of consensus derived from social interaction or communication. This would explain why Factor I acquiescence depends on the presence of plausible generalities. In a sense the subject is conforming to what is generally accepted as truth, just as in social desirability he conforms to some generally acceptable standard of self-description. The conditions surrounding the emergence of Factor I acquiescence make it highly probable that we are dealing with a 'primitive belief' factor such as Rokeach describes in his exposition of the Dogmatism Scale (1960, p. 31). This factor, properly measured, would probably serve as an indicator of readiness to accept 'commonsense' truths. Finally, if the subject finds little guidance in either SD or rational acquiescence, he must fall back on his own tendency to acquiesce or dissent. This is the most ambiguous situation of the three, and here the tendency to dissent may well be a measure of repression.

The theory makes no assertions about the relative power of these factors. Clearly some items and item-clusters strongly evoke one tendency while being neutral for the others. There are individual differences, however, in the tendencies evoked by all three variables and since they do not correlate it is improbable that any great number of individuals is equally affected by two tendencies. The theory would predict that individuals give primacy to one tendency only and that this prime factor would be a major personality variable for the person concerned especially in the area of persuasability and conformity. Thus it would generate the hypothesis that those whose prime factor was rational acquiescence would yield more readily (or react against) statements presented as 'facts' than to other types of persuasion; also the hypothesis that those who accord the primacy to SD would find 'facts' irrelevant if conventionality was simultaneously in question; and thirdly that those who give primacy to the R-factor would include both those who were readiest and those who were most unwilling to act in unstructured situations. The three types of conformity-nonconformity implied by these general hypotheses bear a close resemblance to three types conceptually distinguished by Allport (1962).

## REFERENCES

- ALLPORT, G. W. (1962). Prejudice: Is it societal or personal? *J. Social Issues*, **18**, 2, 120-134.
- BANTA, T. J. (1961). Social Attitudes and response styles. *Educ. psychol. Meas.* **21**, 3, 543-557.
- BASS, B. M. (1956). Development and evaluation of a scale for measuring social acquiescence. *J. abnorm. soc. Psychol.* **53**, 296-299.
- COUCH, A. & KENISTON, K. (1960). Yeasayers and Naysayers: Agreeing Response Set as a Personality Variable. *J. abnorm. soc. Psychol.* **60**, 2, 151-174.
- CRONBACH, L. J. (1946). Response sets and test validity. *Educ. psychol. Meas.* **6**, 475-494.
- CROWNE, D. P. & MARLOWE, D. (1960). A new scale of social desirability independent of pathology. *J. consult. Psychol.* **24**, 4, 349-354.
- EDWARDS, A. L. (1957). *The Social Desirability Variable in Personality Assessment and Research*. New York: The Dryden Press.
- EYSENCK, H. J. (1962). Response set, authoritarianism and personality questionnaires. *Brit. J. soc. clin. Psychol.* **1**, 20-24.
- JACKSON, D. N. & MESSICK, S. (1962). Response styles and the assessment of psychopathology, chap. 8. In *Measurement in Personality and Cognition* (eds. Messick, S., and Ross, J.). New York: John Wiley & Sons.
- KAISER, K. F. (1960). Comments on communalities and the number of factors. Read at an informal conference 'The Communality Problem in Factor Analysis' Washington



- University. St. Louis (dittoed). Cited in Cooley, W. W. & Lohnes, P. R. (1962). *Multivariate Procedures for the Behavioural Sciences*. New York: John Wiley & Sons, Inc.
- LEAVITT, H. J., HAX, H. & ROCHE, J. H. (1955). 'Authoritarianism' and agreement with things authoritative. *J. Psychol.* **40**, 215-221.
- MCGEE, R. K. (1962). The relationship between response style and personality variables: 1. The measurement of response acquiescence. *J. abnorm. soc. Psychol.* **63**, e, 229-233.
- MESSICK, S. (1962). Response style and content measures from personality inventories. *Educ. psychol. Meas.* **22**, 1, 41-56.
- ROKEACH, M. (1960). *The Open and Closed Mind: Investigations into the nature of belief systems and personality systems*. New York: Basic Books.
- RUDIN, S. A. (1961). The Relationship between rational and irrational authoritarianism. *J. Psychol.* **52**, 179-183.
- SROLE, L. (1956). Social integration and certain corollaries: an exploratory study. *Amer. Sociol. Rev.* **21**, 709-716.
- WELSH, G. S. (1956). Factor dimensions A and R. In *Basic Readings on the MMPI in Psychology and Medicine* (eds. Welsh, G. S. & Dahlstrom, W. G.), pp. 264-281. Minneapolis: University of Minnesota Press.
- WIGGINS, J. S. (1962). Definitions of social desirability and acquiescence in personality inventories. Chap 7. In *Measurement in Personality and Cognition* (eds. Messick, S. & Ross, J.). New York: John Wiley & Son.

*Manuscript received 26 April 1963*

*Revised manuscript received 14 May 1963*

## **The Effects of Time of Day and Social Isolation on the Relationship Between Temperament and Performance**

By W. P. COLQUHOUN AND D. W. J. CORCORAN

*Medical Research Council Applied Psychology Research Unit, Cambridge*

In an experiment on cancelling letters in English prose the relationship between output and degree of introversion was found to depend both on the time of day at which the task was performed and on the social situation in which subjects were tested. Performance correlated positively with introversion in morning test sessions when subjects were isolated from each other, but when testing was in the afternoon, or when subjects worked together in a group, this relationship was destroyed. The results are discussed in relation to an arousal theory of introversion-extraversion.

### INTRODUCTION

It has been reported by Colquhoun (1960) that the efficiency with which a subject detects brief, rarely appearing signals when working alone at a paced inspection task is related to his score on the 'unsociability' scale of the Heron personality inventory (Heron, 1956). In tests conducted during the morning good performance was associated with a high unsociability score, i.e. with a high score of introversion; in tests carried out in the afternoon the relationship was reversed. In the experiment to be described this 'time of day' effect was investigated using an unpaced task—letter cancellation—in order to test the generality of the earlier finding. In addition, the effect of social isolation on the relationship was determined by having subjects work either alone, as in the previous study, or, alternatively, when seated together in a single room.

### METHOD

One hundred and twenty-two ratings of the Royal Navy served as subjects. They became available for testing in batches of four, five or six over a period of about 1 year. Each batch of subjects (the members of which were previously unknown to each other) was assigned to one of the following four testing conditions:

1. Morning, isolated (working alone at 8.30 a.m.)
2. Afternoon, isolated (working alone at 1.30 p.m.)
3. Morning, group (working together at 8.30 a.m.)
4. Afternoon, group (working together at 1.30 p.m.)

In the 'isolated' conditions each subject occupied a separate room opening off a corridor from which the experimenter observed their work. In the 'group' conditions the subjects were seated all together around a circular table, while the experimenter remained in the same room.

Each subject was given a pencil and a booklet consisting of several pages of printed English prose. He was asked to check this material line by line for 15 minutes, crossing out each letter 'e' as he did so. The instructions given emphasized both speed and accuracy.

The Heron personality inventory was administered as normal laboratory routine within a few days of the test session.



## RESULTS

*Differences between experimental conditions.* Two scores were obtained: (1) *output*—the number of lines of text checked in 15 minutes; (2) *error*—the percentage of letters 'e' remaining uncanceled. Means and s.d.s of these performance scores and those obtained from the Heron inventory are shown in Table 1.

Table 1. *Scores obtained from a test of cancelling 'e's' in English prose, and from the Heron personality inventory*

Condition	N.	No. of lines checked		Percent omissions		'Unsociability' score	
		Mean	S.D.	Mean	S.D.	Mean	S.D.
1. Morning, isolated	29	85.6	14.3	8.8	6.9		
2. Afternoon, isolated	33	84.0	12.8	7.3	5.0	4.7	2.7
3. Morning, group	30	80.8	10.0	9.7	6.3	3.4	2.2
4. Afternoon, group	30	80.3	11.8	8.0	4.4	4.5	2.4
						4.6	2.9

Although differences in performance between conditions were slight, mean output was consistently greater in the isolated conditions than in the group conditions. The difference of 4.4 lines checked between the overall means in the two situations is just statistically significant at the 5 per cent level of confidence.

Due to chance errors of sampling the mean unsociability score in Condition 2 was significantly lower than the scores in the remaining conditions ( $p < 0.05$ ). However, this is not considered to have affected the conclusions drawn from the main analysis (see below), since when the groups were equated by matching subjects in terms of their unsociability score very similar results were obtained.

*Output  $\times$  introversion.* The relationship between output score and unsociability score in each condition was assessed by product-moment correlation; the coefficients obtained are shown in Table 2.

Table 2. *Product-moment correlations of number of lines checked and introversion score in four testing conditions*

	Morning (M)	Afternoon (A)	Difference (M)—(A)
Isolated (I)	Condition 1 (N 29) +0.443*	Condition 2 (N 33) -0.056	+0.499†
Group (G)	Condition 3 (N 30) -0.164	Condition 4 (N 30) +0.066	-0.230
Difference (I)-(G)	+0.607†	-0.122	

\*  $P(r) < 0.05$  (two-tailed test).

†  $P(\text{Diff.}) < 0.05$  (two-tailed test).

There was a statistically significant positive correlation between unsociability score and output in Condition 1 (the morning, isolated situation). Table 2 shows that this correlation was destroyed either by altering the time of administration of the

test from morning to afternoon, or by testing in a group rather than in isolation. The difference between the correlation coefficients for Condition 1 and Condition 2, and that between Condition 1 and Condition 3 are both statistically significant at the 5 per cent level of confidence.

Table 3 shows the mean output scores of 'extraverted' subjects (those with an

Table 3. *Mean output scores of 'extraverted' and 'introverted' subjects in different testing conditions*

Condition	Extraverted subjects			Introverted subjects		
	N	Mean 'un-sociability' score	No. of lines checked	N	Mean 'un-sociability' score	No. of lines checked
1. Morning, isolated	14	2.4	78.6	15	6.9	92.9
2. Afternoon, isolated	25	2.4	84.4	8	6.4	82.8
3. Morning, group	15	2.4	82.8	15	6.3	78.7
4. Afternoon, group	13	2.0	79.2	17	6.5	81.1

Difference (2)-(1): +5.8  
 Difference (3)-(1): +4.2  
 \*  $t = 2.92, p < 0.01$ .

Difference (2)-(1): -10.1  
 Difference (3)-(1): -14.2\*

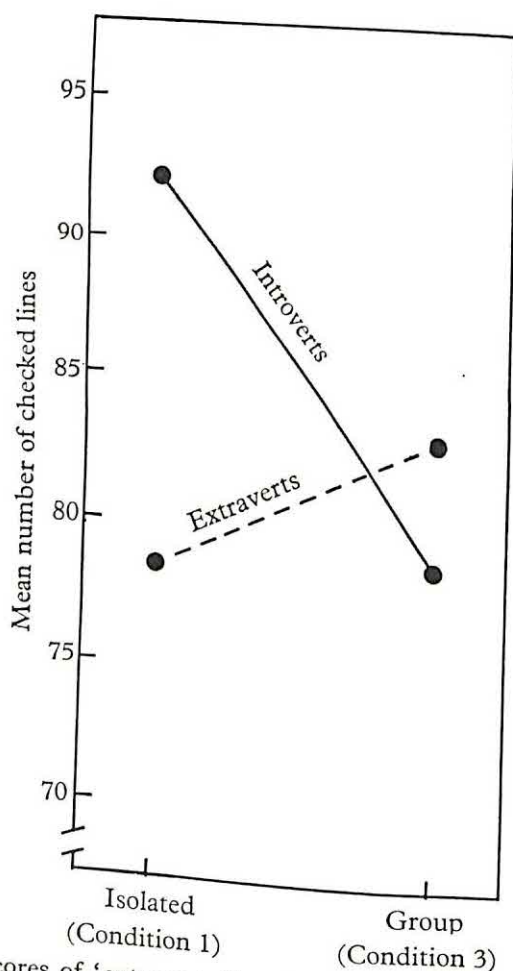


Fig. 1. Mean output scores of 'extraverted' and 'introverted' subjects in different social situations when tested in the morning.



unsociability score of 4 or less) and 'introverted' subjects (those with a score of 5 or more) in each condition. Also shown are the alterations that occurred in the output of extraverts and introverts when the testing conditions of Condition 1 were changed from morning to afternoon, or from the isolated to the group situation.

Fewer lines of text were checked by introverted subjects when tested in a group. The difference is statistically significant ( $p < 0.01$ ). Extraverted subjects

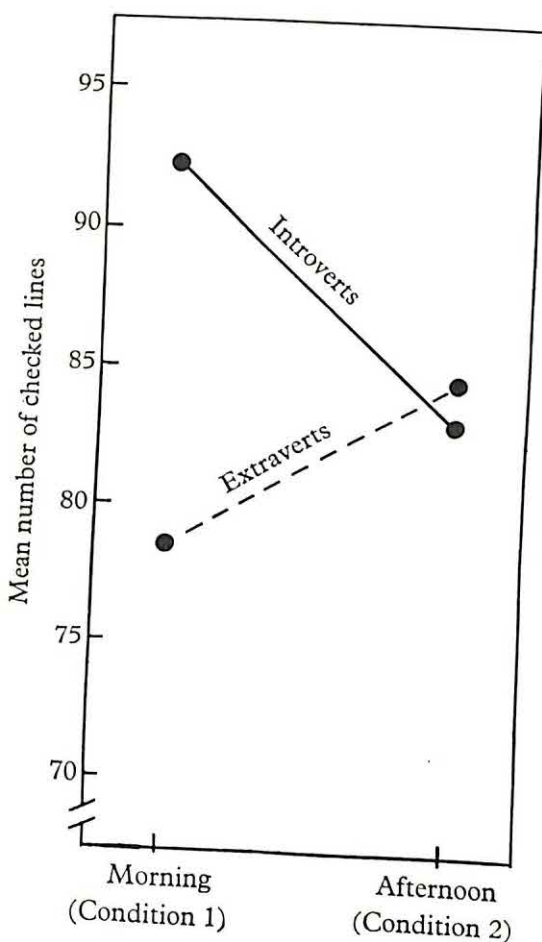


Fig. 2. Mean output scores of 'extraverted' and 'introverted' subjects at different times of day in isolated testing conditions.

tended to increase their output when tested in a group or in the afternoon, but neither of these changes is statistically significant. Finally, introverted subjects tended to reduce their output in the afternoon, but not to a significant extent (Figs. 1 and 2).

**Error.** The pattern of relationships between unsociability score and percentage of letters 'e' remaining uncanceled in the four conditions was similar to that for output, but in this case no individual correlation coefficient, nor any of the differences between pairs of coefficients, were statistically significant (Condition 1,  $r = +0.350$ ; Condition 2,  $r = +0.063$ ; Condition 3,  $r = -0.004$ ; Condition 4,  $r = -0.023$ ).

## DISCUSSION

The positive correlation between output and introversion in the morning, isolated testing condition corroborates the finding of Colquhoun (1960), and demonstrates that this relationship is not specific to the form of inspection task employed. The present results further suggest that the change in the relationship with time of day reported by Colquhoun was due to a differential effect on introverts and extraverts; in the cancellation tasks the performance of introverts would seem to deteriorate and,

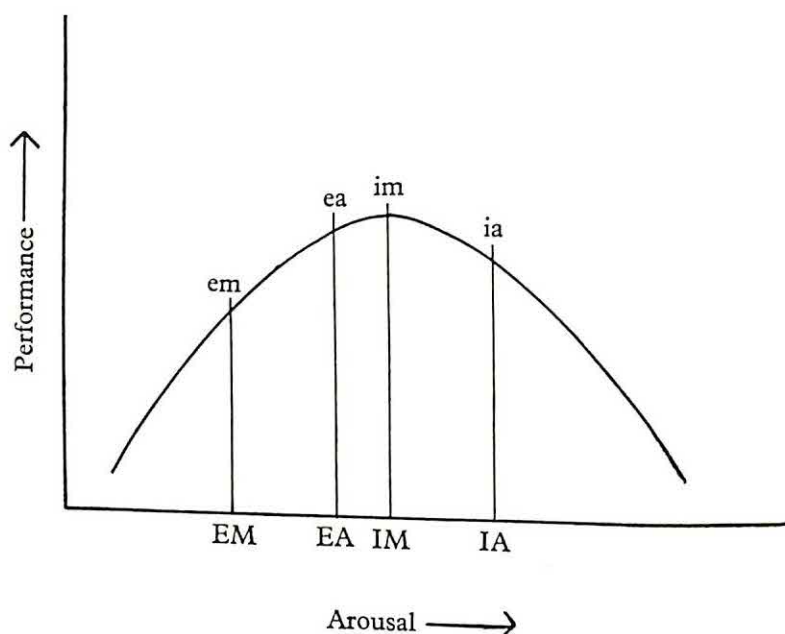


Fig. 3. The 'inverted-U' relationship between performance and arousal used as a possible explanation of the results (see text).

to a lesser extent, that of extraverts to improve when testing takes place in the afternoon rather than in the morning. Similar effects were observed on changing from isolated to group testing conditions.

Although these results are difficult to explain in terms of existing theories (e.g. Eysenck, 1957; Spence, 1957), it is possible to account for the findings in terms of a model for the effects of incentives and hyper-arousing situations suggested by Corcoran (1962). This assumes that (a) introverts are characteristically higher in arousal than extraverts, (b) that the relationship between level of performance and degree of arousal is inverted-U shaped, and (c) that certain environmental and task conditions affect level of arousal. The inverted-U relationship is shown diagrammatically in Fig. 3.

Since such measures as body temperature and heart rate indicate that level of arousal is low in the morning (Kleitman, 1939), the first assumption of the model (a) would place introverts in a more favourable (optimal) position at this time, say IM in Fig. 3. Extraverts, on the other hand, would be at some position to the left of IM, say EM. A general rise in arousal in the afternoon would shift EM to EA and IM to IA, which would result in poorer performance by introverts and better



performance by extraverts. The same kind of model can be applied to the change from isolated to group testing conditions, provided it is assumed that the incentive provided by group testing raises level of arousal.

Finally, it should be emphasized that separate groups of subjects were used under the different conditions, and that although a definite alteration in the relationship between introversion-extraversion and performance has been shown, the precise effect of the changes in testing conditions on individual subjects of a particular temperamental type requires further investigation.

Thanks are due to the Royal Navy for supplying the subjects for these experiments, and to Mr P. Freeman for statistical advice.

#### REFERENCES

- COLQUHOUN, W. P. (1960). Temperament, inspection efficiency and time of day. *Ergonomics*, 3, 377-378.
- CORCORAN, D. W. J. (1962). Individual differences in performance after loss of sleep. Unpublished Ph.D. Thesis, University of Cambridge.
- EYSENCK, H. J. (1957). *The Dynamics of Anxiety and Hysteria*. London: Routledge and Kegan Paul.
- HERON, A. (1956). A two part personality inventory for use as a research criterion. *Brit. J. Psychol.* 47, 243-251.
- KLEITMAN, N. (1939). *Sleep and Wakefulness*. Chicago: University Press.
- SPENCE, K. W. (1957). *Behaviour Theory and Conditioning*. Yale: University Press.

*Manuscript received 25 June 1963*

*Revised manuscript received 26 October 1963*

## Book Reviews

*Personality Development and Psychopathology.* By Norman Cameron. Houghton Mifflin Company, Boston, 1963. Pp. 793. \$8.75.

Norman Cameron is an eminent psychiatrist and well known as an exponent of Adolf Meyer's biosocial approach. Whilst not rejecting this, he has now moved across to a psychodynamic viewpoint which leans heavily on psychoanalysis. This is, in fact, two books in one, since development psychology is covered in the early chapters and leads on, via an outline of Freudian theory, to a treatment of psychopathology. Some topics receive short thrift and often alternative viewpoints are mentioned only briefly. Nevertheless the references are up-to-date and useful, and such relevant topics as sensory deprivation and the experimental work on overinclusiveness are included. Psychoanalytical terms abound and the sceptic will find much to be sceptical about. Dr Cameron states, however, that he is especially concerned with the 'inner life of man', and the strong point of the book is the author's clinical experience.

PETER MAYO

*Clinical Psychology: An Introduction to Research and Practice.* By N. D. Sundberg and L. E. Tyler. London: Methuen, 1963. Pp. xviii + 564. 50s.

This book is not in any way a text-book or manual of the techniques of clinical psychology. Actual testing procedures are relegated to an appendix of 'fifty tests of importance in clinical psychology', each with a very brief description. It is really an exhaustive examination of the 'philosophy' of clinical psychology, its aims and the methods of achieving them.

The two main sections, together making up three-quarters of the book, are on 'Psychological assessment' and 'Psychological approaches to treatment'. The section on assessment is a detailed consideration of the problems involved in making meaningful and useful statements about people, the contribution of interviews, tests, observation and life histories being carefully evaluated. The treatment section consists of a brief description of a great many methods of psychotherapy classified by an idiosyncratic method (but probably as good as any other!) under seven chapter headings—'Psychotherapy as exploration of resources', 'Psychotherapy as habit change', etc. Each chapter ends with a look at attempts which have been made to assess the effectiveness of each type of therapy—and sad reading it makes. To quote one of the few jokes in the book (p. 420): '... the publications that result boil down to little more than this: "We tried group therapy with wayward girls and we liked it"'.

The authors' approach is throughout critical and committed to nothing except the patient's welfare. On all topics pros are well balanced by cons. Statements are well documented, and each chapter is followed by summaries of relevant research papers. The book certainly leaves you with the impression that clinical psychology contains a great excess of problems over solutions, but then so it does, and it is a tribute to the authors' objectiveness that no other impression is given. A more slanted approach would probably have made easier reading, but the book is written primarily for students and the aim has been to present them with a panoramic view of the methods and theories of clinical psychologists. One feels that an inexperienced student might be discouraged by the multiplicity of approaches exposed, and that the book might perhaps be more beneficial to the older practising psychologist as an antidote to the hardening of professional arteries.

A chauvinistic Briton could object to the All-American nature of this book. Foulds' work on diagnosis, for instance, is not mentioned; Eysenck's study of the effectiveness of psychotherapy is discussed, but his whole personality theory is only mentioned *en passant*. Psychological contributions in the neurological field receive little attention. However, to have summarized so much of American thinking on this complex subject must qualify as a considerable achievement.

T. G. CROOKES



*Pickford Projective Pictures.* By R. W. Pickford *et al.* Tavistock Publications, 1963. Text: 30s. Picture material: 25s. The set: 50s.

This new material consists of 120 (*sic*) postcard-size drawings, done in impressionistic black outline on a white ground, which depict '... one or more children in as large a variety of everyday situations as possible'; such situations often involving adult figures and even the occasional fantastic animal. The pictures are intended mainly, but not exclusively, for use with children, and the procedure has been evolved very largely in a child-guidance setting. The companion book contains a certain amount of 'normative' data about the range of perceptual and apperceptual interpretations which children have hitherto tended to put upon the stimulus-material (the relevant section having been compiled by John S. Struthers): it also discusses various ways of using the pictures, and presents case studies (some of them by Miss L. R. Bowyer).

One may ask whether we really need more projective material that is not geared to a specific technique of testing or to a thoroughgoing rationale of perceptual dynamics, as is, for example, Raven's or Phillipson's. After all, if all the projective tests that already exist were put end to end, they would stretch from the Collective Unconscious to the Superego. Professor Pickford would claim, however, that his set has, quite apart from the questionable merit of inexhaustibility, some novel applications.

For it is intended as a *therapeutic* as much as (if not more than) a diagnostic instrument: that is to say, as a means by which the disturbed child not merely communicates his conflicts etc. to someone, but actually works through them *within* an on-going interpersonal (perhaps a 'transference-' relationship. But the writer insists that there is no stipulated way of achieving this end. The therapist may either make interpretive interventions or sit in solemn silence, according to his own lights. Indeed, so apparently concerned is the author to avoid dictating structure to either the patient or the psychologist (lest he thereby imply some particular psychodynamic allegiance?) that he is to be found saying at one point (p. 4) that his pictures 'are not usually aimed in any definite way at particular fantasies . . .', although he later (p. 7) classifies all the cards into twenty groups 'according to content, subject-matter or theme . . .', under such titles as 'sibling jealousy' and 'conflict with father figure'. A different set of twenty batches, each of varied content this time, is also suggested, and these could be used, one group per session, for the sort of extended period of therapeutic communication that the author has in mind.

This latter use will be recognized by most clinicians who attempt child-therapy as the sort of device they have often wanted for the verbal, but inhibited or schizoid, child who finds it too effortful or threatening to graduate from communicating feeling and fantasies *via* picture-stories to expressing them directly to his therapist. It is, of course, possible to use an extended 'inquiry' to, for example, T.A.T. or O.R.T. responses in a similar way with older children and adults, and therefore in saying that other projective methods are 'essentially diagnostic' Professor Pickford perhaps underestimates their versatility. But the new pictures are, no doubt, uniquely suitable for this sort of purpose.

It must be added that the so-called 'normative' data are derived from a clinic population; that for twenty of the cards  $n$  is less than 12; and that, considering girls only,  $n$  is less than 10 for almost half the material. This being so, the author's claim that 'by reference to these data the approximate degree of abnormality in a given child's responses may be estimated in terms of frequencies' is misleadingly extravagant. Nor will those sceptical of the value of projective techniques as a whole find any comfort in the author's generally rather tender-minded and unrigorous remarks about the nature of projective and psychotherapeutic processes, in course of which the role of the passively accepting listener is repeatedly stressed. What is one to make, for instance, of this sort of assertion (p. 11): 'the present writer . . . has treated many children successfully without using verbal interpretation at all, and it is not likely that the therapy in these cases could have been cut shorter by any other method'? However, as is the case with many a psychodynamic enterprise, the fantasy-material (or 'verbal behaviour', if you must) which it elicits may be thought more persuasive than what is said about the enterprise itself.



*Social Learning and Personality Development.* By Albert Bandura & Richard H. Walters. New York: Holt, Rinehart and Winston, 1963. Pp. 329. \$7.00.

In this new text principles of learning provide the framework for the study of personality development. The approach is in terms of learning theory; but it is learning theory with a difference, for in contrast to some other formulations with which it has kinship, the volume consistently and systematically emphasizes the role of social variables in shaping the individual personality. The authors' main aim is 'to explain the development of all forms of social behavior in terms of antecedent social stimulus events such as the behavioral characteristics of the social models to which a child has been exposed, the reinforcement contingencies of his learning history, and the methods of training that have been used to develop and modify his social behavior' (p. 44). A special merit of the viewpoint presented is the emphasis that one can account for socially acceptable and unacceptable behaviour alike by means of the same principles of learning; the acquisition and modification of prosocial and deviant behaviour can be examined through the same conceptual glasses.

Throughout the volume the authors demonstrate mastery of the experimental literature, and they buttress their position by drawing on data from the observations of anthropologists, sociologists, ethologists, social psychologists and clinicians. For despite the fact that the authors are hard-shell experimentalists, they realistically note some of the limitations of experimentation and point out the important role that other methods may play as part of a sound research strategy for discovering principles of personality development.

After an introductory statement outlining the general nature of the approach, subsequent chapters deal extensively with the role of imitation, reinforcement patterns in the development of behaviour, the acquisition and maintenance of self-control, and a specification of procedures that might be used in the modification of behaviour.

The attention given to the role of imitation is clearly consistent with an approach emphasizing the contribution of social variables. The authors note three significant ways that the behaviour of models may influence the learner: (a) a modelling effect (the transmission of precisely imitated response patterns that were not previously a part of the observer's repertoire), (b) an inhibitory or disinhibitory effect on previously acquired observer responses that are similar to those exhibited by the model, and (c) an eliciting effect, in which observer responses that are neither entirely novel nor inhibited as a result of prior learning are 'released' by behavioural cues provided by the model. This treatment of imitation goes well beyond a conceptualization of it as a form of instrumental conditioning; the authors explicitly consider as well the vicarious reinforcement stemming from response consequences to the model. The observer's knowledge of such consequences emerges as a relevant element in determining the nature of his own subsequent actions.

A reappraisal and revision of the frustration-aggression hypothesis is an integral part of a discussion devoted to the role of reinforcement patterns in establishing, maintaining, or inhibiting social behaviour: Frustration is not viewed as an inevitable antecedent of aggression, for 'one can readily produce a highly aggressive child by merely exposing him to successful aggressive models and rewarding the child intermittently for aggressive behaviour while keeping frustration at a very low level' (p. 159). And aggression is not necessarily the inevitable, or even the naturally dominant response to frustration, since the nature of the response 'will depend on the prior social training of the frustrated subject . . .' (p. 159).

The conceptual scheme for social learning that is erected is a complicated structure. But the reader's task is simplified by the manner in which each component of the structure is held up to view. Thus in a chapter on the development of self-control the authors do a commendable job not only in presenting information on the operation of such factors as modeling, direct reinforcement and disciplinary techniques; they also point to the way these and other elements are entwined. For example, they note the interplay of the constitutional or acquired characteristics of the learner, his prior social-learning experiences, the context in which reinforcement occurs, the intensity, timing and sequence of techniques intended to shape behaviour, the effects of particular reinforcement combinations, and the inconsistent actions and unwitting provision of cues by role models and disciplinary agents. Such a specification of factors entering into the social-learning equation does not, of course, yield a quick and ready solution to problems of how best to guide personality development. But it does serve to con-



tribute guide-lines useful to the research worker interested in more precise specification of the learning process and should prove helpful to the practitioner as an aid in sensitizing him to the potentials in the broad spectrum of factors relevant to the modification of behaviour.

To the extent that the book has a villain (or at least a whipping boy) the authors have reserved this role for most of the traditional psychodynamic theories of personality. Although there is much that is sound in their appraisal of such theories and the therapy extending from them, the critique contains few novel points and is in some respects tangential to the book's aim of presenting a socio-behaviouristic approach to personality development.

This volume is a relevant and valuable contribution to our knowledge of personality development and social learning. It coherently presents a systematic and carefully documented point of view. As such it should stimulate interest among students, and for practitioners and research workers it should evoke not only some controversy, but also an additional measure of fruitful inquiry.

JOHN D. CAMPBELL

*Stress and Release in an Urban Estate—A Study in Action Research.* By John Spencer, with the collaboration of Joy Tuxford and Norman Dennis. London: Tavistock Publications, 1964. Pp. xiv + 355. 45s.

In 1952 a group of citizens in Bristol, concerned about delinquency and other social disturbances, put up a proposal for a project involving both social work and research. Some years earlier Mr John Mack had suggested something very similar in a report on delinquency to the Carnegie Foundation, and in 1953 the Foundation set up the Bristol Social Project to carry through a programme on these lines. Mr John Spencer directed the project and he had the assistance of eleven other people of varying periods.

*Stress and Release in an Urban Estate* is the first of two volumes giving an account of the project. Almost two-thirds of the book is devoted to descriptions and analysis of various enterprises which the project team was concerned with in one way and another. One of these enterprises, for example, was the initiation and conduct of meetings for a group of socially ineffective housewives, another the running of a group of difficult adolescents and another the running of something like a seminar for a group of teachers. These accounts provide excellent case material against which club leaders and others can match their own experiences and they will be useful as teaching material for social work students intending to enter group work.

For the academic reader with no special concern for social work the main interest of the book will be in the introductory and concluding sections in which the authors describe their special 'action-research' approach. They stress that this approach is in 'marked contrast to the conventional methods of social research'. As I understand it, it differs from conventional research in that it restricts itself, at least in the early stages, to problems defined and felt as such by lay people as opposed to professional social scientists. Further, it deliberately attempts to make the people studied more sophisticated about themselves and others and to help them to achieve certain goals. It is not clear what happens if research as such clashes with these additional requirements, but the impression is that in such a circumstance it is the research which gives way.

Reading this as sympathetically as possible I fail to see how anyone can be convinced that the authors were engaged in a type of activity significantly different either from social work or research as conventionally understood. And their conviction that they were has clearly been a handicap, for it has led them to dismiss both the literature and the common experience of social research as irrelevant to their problems. The observations they make (p. 111), for example, on the effects of hierarchy on social interaction are the common currency of the undergraduate courses on social stratification which at some time they must have taken. It is depressing to find the lessons so laboriously and expensively re-learned. The difficulties which the project ran into with its steering committee and other bodies seem also to be much like those met with in most large social research projects and cannot really be attributed to the special nature of the work although, admittedly, such difficulties are not usually given such a frank and piquant discussion as in *Stress and Release in an Urban Estate*.

Dr Spencer and his collaborators have made a contribution to the literature of social work but they have done nothing to establish 'action-research' as a viable method of social enquiry.

PETER COLLINS



*Integrating Principles of Social Psychology.* By J. B. Cooper & J. L. McGaugh. Cambridge, Mass.: Schenkman, 1963. Pp. xi-320. \$7.50.

This book is intended as a general text in social psychology for students. It is written from a point of view which has been somewhat neglected in recent years, and is welcome for that reason alone. The defining characteristics of this point of view are three: an emphasis upon understanding social behaviour in terms of the individual—the authors explicitly reject an approach which takes as the basic units of analysis the two-person or group situation; secondly, what the authors call a 'cognitive-theoretical' position—this is never made explicit, but the ghost of Lewin is present in much of the book; and lastly, and most important of all, a strong interest in the comparative study of social behaviour.

Plainly, these 'integrating principles' do not integrate social psychology, but they do give coherence to the book by providing a basis for selection and emphasis. Especially is this true of the third 'principle'. The comparative approach is adopted wherever possible, even, for instance, in the chapter on prejudice. It is responsible for some of the best parts of the book, such as the accounts of language and communication, and leadership. However, no attempt is made to treat any topic in a thoroughly systematic and comprehensive fashion, and some chapters are a long way from this ideal. Thus, a curious chapter which purports to deal with the 'problem of human nature' contains, in order, very brief summaries of the views of some famous thinkers, a defence of phenomenology together with an exposition of Lewin's adoption of the genotype-phenotype distinction, and ends with an account of the semantic differential. The chapter on learning and socialization contains an account of learning which is little more than a gesture and omits any treatment of imitation, though it does, on the other hand, include some detailed reference to animal studies of the effects of early experience. There is a great deal of empirical reference in the book, but it is unevenly distributed, some topics receive virtually none and others getting excessively detailed treatment.

In short, the book is too variable in style and content to become a standard text for students, but it should serve as useful supplementary reading for those interested in the comparative approach.

DEREK WRIGHT

*On the Theory of Social Change.* By Everett E. Hagen. London: Tavistock, 1964. Pp. xvii + 502 + Appendices, Bibliography and Index. 63s.

Hagen writes as an economist critical of the theories of economic growth and social change which have been advanced and accepted in the field of economics. Technological change, rather than mere capital formation, he sees as basic to both increase in productivity and to the re-structuring of social relationships. Technological change can occur only as there are changes in social organization to make use of such changes. This implies the development of new value-structures and patterns of motivation. Hagen is not, however, a technological determinist—far from it. New techniques in themselves are adapted only by certain personality types since even imitation implies as certain creativity. He is thus concerned with what he calls innovational personalities and their location in the social structure, as the sponsors of change; and, less directly with authoritarian personalities who resist change. His innovators are located in socially disinherited groups (not entirely identifiable with pariah groups) among whom, after a period of retreatist response to the wider society, innovators arise as a product of the interaction of particular personality types among parents and children within the group. His psychological discussion derives its insights largely from Erikson.

Hagen is thus concerned with a more complex analysis than that accepted by either cultural diffusionists or sociological writers on social change. He is not merely concerned with transmission of techniques, nor merely with change of value structures which facilitate their adoption, but also with the development of innovating personalities. He seeks to illustrate his thesis with case studies of Norman England, Japan, Columbia, Java, Burma and on the Sioux Indian reservation.

BRYAN WILSON



*Trends in the Mental Health Services.* Edited by Hugh Freeman and James Farndale  
London: Pergamon Press, 1963. Pp. xv + 341. 70s.

This book is a collection of papers concerned with organizational and administrative problems in the mental health services. Some have already been published, but many are printed here for the first time. Although some degree of overlap and well-meaning platitudinizing are probably unavoidable in a volume of this kind, the standard of the papers is generally high, and they are for the most part informative and readable. The book is really indispensable for anyone wishing to familiarize himself with present-day thinking on reorganization and planning of psychiatric services in hospital and community settings.

The largest number of papers are devoted to hospital problems, and some of these present the well-known statistical forecasts of patient trends on which national planning is being based, together with some of the objections and controversies which they provoked. It is most useful to have these papers brought together, so that the argument can be followed more closely. There are several accounts of experimental schemes, in particular the comprehensive service at Nottingham which was, in fact, in existence before 1948; the industrial therapy projects at Bristol and Cheadle Royal; and a careful study of the effects of the Worthing scheme. There is a long account from Birmingham of the plans and problems of a 'Balanced Hospital Community'. By no means the least profound comes in the form of a delightful *jeu d'esprit* by Maddison, solemnly entitled, 'Blueprint for a model psychiatric hospital of the future', in which he uses the imagined occasion of the opening of a new hospital in 1999 to survey the extraordinary attitudes and habits of psychiatrists and administrators in the first half of the twentieth century. The attack is devastating and the satire vicious.

If anyone still doubts that the running of a mental hospital or a community service needs outstanding administrative ability or supposes that any reasonably competent psychiatrist can get through his administrative chores with a few letters and committee meetings, this book should provide a salutary corrective. Rehin and Martin, in an excellent account of research problems in community care, go so far as to suggest that 'perhaps mental health is too serious and important a matter for policy decisions to be left to psychiatrists'. Although some will be shocked by such a heresy, there have for some time been ominous rumours that perhaps the qualities that make for a good psychiatrist are not necessarily those that make for the good administrator. One day we may even see a job analysis of the work of a medical superintendent.

The book provides many examples of serious, closely reasoned argument about the problems posed by the current concern with community care and psychiatric wards in general hospitals. Kathleen Jones, for example, argues strongly that the general hospital by its very nature cannot hope to provide adequately for psychiatric patients, far less replace the mental hospital. 'Only in Britain', she concludes, 'has progress taken the form of denying that the mental hospital has a useful function, and planning for its abolition.' Other contributors provide accounts of modernization schemes within hospitals in terms not only of buildings and amenities, but of attempts to provide a therapeutic social climate. To do this requires a good deal of rethinking of the part played by nurses, and this is admirably demonstrated by Annie Altschul.

After reading this book one is left with the conviction that research must receive much greater support in future. It is extraordinary that far-reaching and quite progressive planning at a national level can be built up on such scanty information as is currently available, and the inspired guesswork that now passes for planning will, one hopes, be replaced in the future by decisions based on properly conducted and generously financed sociological and epidemiological research.

In the meantime, unless progress is much more rapid than most of its contributors foresee, this book should remain a standard work of reference for some time to come, and the editors are to be congratulated on their achievement.



*Social Pressures in Informal Groups: a Study of Human Factors in Housing.* By L. Festinger, S. Schachter and K. Back. London: Tavistock Publications, 1963. Pp. x+197. 38s.

This is a re-issue of the well-known study which first appeared in 1950. *Social Pressures in Informal Groups* has been of value to many people, from those primarily concerned with the social psychology of housing to those more interested in questions of theory and methodology. Certainly, some of it still seems rather trivial and obvious. The reviewer has never liked the language of 'valences' and 'force fields' and indeed tends to find it quite confusing at times; and he continues to have doubts about the general usefulness of the concept of 'cohesiveness'. Nevertheless, there is a lot here that is still worth reading and thinking about, and this re-issue is very welcome. It does, however, appear rather expensive at the price.

D. GRAHAM

*Probation and Mental Treatment.* By M. Grünhut. London: Library of Criminology No. 10, Tavistock Publications, 1963. Pp. vii+56. 15s.

Superficially, this little book makes dull reading, but closer inspection reveals innumerable hidden jewels. This research from the University of Oxford is based on a 1953 sample of offenders who were dealt with by the courts under section 4 of the Criminal Justice Act, 1948, requiring the offender to undergo medical treatment in conjunction with probation for up to 12 months. Out of 882 cases, 636 were investigated of which 275 offences were committed against property, 216 were sexual, 65 suicides, 32 violence against others, and 48 were miscellaneous, like loitering or being drunk. The classification of medical reports for 414 probationers required to submit to mental treatment showed that 195 were classed as psychopaths, 28 low intelligence, 58 schizophrenics, 52 depressions, 38 neurotics and anxiety states, 36 illness with a physical basis. As expected, the majority were treated by psychotherapy, exhibitionists and thieves having a good prognosis, the dishonest and psychopaths a poor one, often depending on the background of the offenders. Briefly, the encouraging picture emerged that out of 100 offenders who receive mental treatment, 70 are discharged with a favourable prognosis, out of these 70, 56 are not reconvicted within 12 months after termination of probation, and a further 13 out of 30 with an unfavourable medical prognosis are also non-recidivists. It is to be hoped that there will not be another time-lag of 10 years before the results will be known of applying the more recent section 60 of the Mental Health Act, 1959 according to which more flexible 'guardianship' treatment orders can be made by comparison with section 4 of the Criminal Justice Act 1948.

R. G. ANDRY

*Emotional Flexibility-Rigidity as a Comprehensive Dimension of Mind.* By Sigvard Rubenowitz. Swedish Council for Personnel Administration. Stockholm: Alnquist, Wiksell, 1963. Pp. 256. 4 Swedish Crowns.

This Swedish report states that its aim is to find which factors make people flexible, and applying this information to industry.

It turns out to be a model for scientific investigation into personality, with clearly stated hypothesis, methods and tests (though the latter have had to be standardized Swedish versions of American tests). Numbers are adequate (3000 have participated) and a factor analysis used to study the dimensionality of the scales.

Evidence was found for a general factor of flexibility-rigidity and a wide variety of interesting hypothesis tested. One of the more important suggestions is that cultural factors cause variations in rigidity and that early family life is a most important determinant in the rigidity-flexibility continuum.

A. W. LLOYD



*A Study of Brief Psychotherapy.* By D. H. Malan. Monographs, No. 8. London: Tavistock Publications, Charles C. Thomas. Pp. 312. 35s.

There are several valuable lessons to be learnt from this thoughtful account of twenty-one out-patient therapies conducted at the Tavistock Clinic. Psychotherapy is an art which has eluded scientific validation, but Dr Malan shows how the problem must be approached if a more meaningful attempt is to be made. While he may not have found all the answers he has provided a useful corrective to the Eysenckian viewpoint. He urges a dynamic rather than merely symptomatic assessment of the presenting problem, with important implications for prognosis and follow-up. At a more practical level he is a persuasive advocate of short-term 'focal' psychotherapy (usually between 10-40 sessions) in patients who are reluctant or cannot afford to undergo years of psychoanalysis. It remains uncertain how such therapy achieves its results, but enthusiasm is regarded as a possibly crucial factor in both parties. While the author's clinical evidence may carry more conviction than his statistical analysis of a small sample, he will command respect from all who care about the rationale as well as the practice of psychotherapy.

MICHAEL HUMPHREY

*Alcoholic Addiction. A Psycho-social Approach to Abnormal Drinking.* By Howard Jones. London: Tavistock Publications, 1963. Pp. x+209. 30s.

This book is an account of a study of alcoholics conducted by the author in Toronto. Seventy-two men, half of them in prison for being drunk in a public place, the rest current cases in the residential treatment unit of the Ontario Alcoholism Research Foundation, were interviewed by twelve graduate social-work students. The interviews were structured on a schedule which was 'not to be followed religiously . . . but to serve as a guide . . . and as an instrument for recording'.

The author has applied the life-history approach as the methodological basis of his study. This precluded specifically defined hypotheses, but not those of a more general character. The claim is made that an overall pattern emerges showing the motivation for addiction. Excessive drinking in the Anglo-Saxon culture is considered the means for the solution of deeper-lying problems. Early maternal dependence and early emotional deprivation constitute the basic dichotomy of problems which lead to the development of 'alcoholic personality' based on overdependence or social insecurity. Seven main 'states of mind' leading to addiction are distinguished: maternal dependence, ego-need, social inadequacy, social dependence, escapism, homosexuality and a latent oedipal fixation. This list is not exhaustive.

Jones is fully aware of the non-representative character of his sample and of its small size (in the chi-square tests many cells have a frequency below five). Yet he proceeds to the statistical analysis of his data and significance tests of some of the hypotheses. This gives the wrong impression that after all the warnings given by the author some dependable results were obtained after all. This is not the case especially as in spite of the far-reaching differences between the two sub-groups (Chapter 13) they were combined for the purpose of analysis (Tables 5-13). In these circumstances the statistical analysis leaves one confused. If this is disregarded there is still plenty of interesting material left. The data concerning the 'contribution' of the wife to the husband's addiction (the domineering wife needing a dependent husband unconsciously encourages his drinking) are worth following-up. The same is true of the proposition that alcoholism is more often a replacement of crime (for unsuccessful criminals) than its cause. An interesting book, if one is careful about the conclusions.

BERNARD MANDELL

*Contributions to Modern Psychology: Selected Readings in General Psychology.* Edited by D. E. Dulany, R. L. DeValois, D. C. Beardslee and M. R. Winterbottom. New York and London: Oxford University Press, 1963. Pp. xii + 484. 28s.

This is the second, revised and enlarged edition, published as a paperback, of the volume of readings which originally appeared in 1958. The book sets out to provide the student with an easy access to some outstanding research reports that have appeared in technical journals, for the most part during the last 30 years. There are nine sections, each containing from four to seven papers, as follows: I—Development, II—Intelligence and Individual Differences, III—Perception, IV—Learning and Memory, V—Thinking and Imagination, VI—Motivation and Emotion, VII—Personality, VIII—Personality Disorders and their Treatment, IX—Social Psychology. The last section on social psychology is a new feature of this edition. Otherwise, some papers have been left out of the present edition and others have been added throughout the volume. The effect of these changes has been that the present book is longer by about a quarter than the one it supersedes. This book is good value for money. The original articles have been well selected, and each is prefaced by a short informative editorial note. The book will be useful to students of psychology in supplementing general text-books; and it will be useful to teachers of general psychology by providing them with a wealth of source material within the covers of a single volume.

W. SLUCKIN

*The Rorschach in Practice.* By Theodora Alcock. London: Tavistock Publications, 1963. Pp. xii + 252. 63s.

The book consists of two main sections. Part I describes (and illustrates) the test itself and methods of scoring the responses to it, while Part II consists of illustrative cases and their interpretations. A short article is appended as Part III demonstrating the use of the Rorschach as a research tool in the investigation of the Asthmatic Personality.

Part II is introduced by a useful discussion on the nature of psychological 'normality' and adjustment. The author herself interprets most of the abnormalities in the patients' records in psychoanalytic terminology, but even non-analytically orientated clinical psychologists should find this a useful reference book.

MOYRA WILLIAMS







